The Neural Basis of the Number Sense

by

Nicholas Kurshan DeWind

Department of Neurobiology
Duke University

Date:_____
Approved:

_____
Michael Platt, Supervisor

_____
Elizabeth Brannon, Chair

_____
Jennifer Groh

_____
Katherine Heller

_____
Marc Sommer

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Neurobiology in the Graduate School
of Duke University

2014

ABSTRACT

The Neural Basis of the Number Sense

by

Nicholas Kurshan DeWind

Department of Neurobiology
Duke University

Date:_____

Approved:

_____
Michael Platt, Supervisor

_____
Elizabeth Brannon, Chair

_____
Jennifer Groh

_____
Katherine Heller

_____
Marc Sommer

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Neurobiology in the Graduate School of Duke University

2014

Abstract

The ability to enumerate approximately without counting is an evolutionarily ancient and developmentally early core cognitive ability known as the “number sense”. We use the number sense when we estimate a number without counting individual items, as when we guess the number of people in a crowded room. The number sense is theorized to form an instinctual building block upon which we create the conceptual structures of mathematics. This dissertation addresses three research questions regarding the number sense.

The first is the question of whether the number sense is malleable, and if so, what are the neural correlates of malleability. In Chapter 2 we gave adults number sense training, which we found improved the accuracy of numerical estimation. In Chapter 4 we recorded from single neurons in monkeys while they viewed arrays of items on a computer screen. Similar recordings have been made previously, but usually using monkeys that were trained to discriminate sets based on number. Recordings in trained animals demonstrated that individual neurons in the monkey’s brain track the number of items in a set. We reasoned that if the neural correlates of the number sense were altered by the training experience, then we would get different results in untrained monkeys. We did find neurons encoding numerical information in untrained monkeys, but at lower rates than described previously. Thus, we demonstrated that the number

sense can improve with experience, and our data suggest that changes in the proportion of neurons encoding number may subserve this improvement.

The second question is how to resolve the problem of stimulus control in laboratory tests of the number sense. Typically, number sense function is assessed by presenting arrays of dots on a computer screen. In such stimuli, however, non-numerical features necessarily covary with numerical features. By counter-balancing different stimulus conditions, it is possible to determine if number and not some other feature is influencing a dependent measure. In Chapter 3, we develop a technique to go further and determine which of eleven stimulus features is influencing a dependent measure.

The third question is whether the intraparietal sulcus (IPS), a brain area known to be engaged during numerical cognition, is specialized for it. To address this question, we apply the technique developed in Chapter 3 to the neural data recorded from monkeys in Chapter 4. We show that the IPS does contain number neurons; however, it also contains neurons that encode many other features in equal proportion, indicating that it is not specialized for number. In Chapter 5, we use drugs injected into the IPS to reversibly inactivate it. We found that after IPS inactivation, performance on a numerical discrimination task was impaired but no more so than a color discrimination control task. Again, our data do not support the theory that the IPS is specialized for numerical processing.

Dedication

In loving memory of Steve Whitman, a great person and a cool dude. I miss him very much.

Contents

List of Tables

List of Figures

List of Abbreviations

AIP	anterior intraparietal area
ANS	approximate number system
AMS	approximate magnitude system
ATOM	a theory of magnitude
Close	apparent closeness
COV	coefficient of variation
Cover	coverage
DLPFC	dorsolateral prefrontal cortex
FA	field area
fMRI	functional magnetic resonance imaging
hIPS	horizontal segiment of the intraparietal sulcus
IP	item perimeter
IPS	intraparietal sulcus
ISA	item surface area
LIP	lateral intraparietal area
Num	numerosity or number
PFC	prefrontal cortex
PPC	posterior parietal cortex
Spac/Sp	*Spacing*
Spar	sparsity
TP	total perimeter
TSA	total surface area
VIP	ventral intraparietal area

Acknowledgements

First and foremost, I want to acknowledge my advisors Michael Platt and Liz Brannon. They provided the expertise and background necessary to design and execute the experiments and taught me how to write a scientific manuscript. More than any other individuals they are responsible for my transformation from lowly first-year to scientist.

I would also like to acknowledge the other coauthors of the manuscripts on which the research chapters are based. The stimulus space and regression approach described in Chapter 3 and 4 was the product of many long conversations with Geoffrey Adams. If Geoff was not so generous with his time and intellectual capacity, I doubt that nut would have cracked. Dr. Jiyun Peng and Andrew Luo both helped collect the data presented in Chapter 5. The undergraduate research assistants in Dr. Brannon's lab helped to collect the data presented in Chapters 2 and 3.

I'm also indebted to all my family and friends who supported me in this endeavor. Most especially Monica Carlson, who has been there for me through thick and thin. She is the glue that holds Dr. Platt's lab together and the best lab manager there is. She has dutifully served as my editor and therapist as required. She is also my very best friend, and I love her very much.

I also want to acknowledge the members of Platt and Brannon labs past and present, with a special shout-out to Arani Roy, who taught me electrophysiology. Together we maintained a rigorous, stimulating, and fun, if not always very professional, intellectual environment. We have always been a good team and I'm very grateful for their wisdom and their company.

I'm also very appreciative to the funding agencies that make it possible. Broad Research Award for Graduate Students to me, McDonnell Scholar Award to Dr. Brannon, and National Eye Institute 1R01EY01 to Dr. Platt all supported the research here.

1. Introduction

1.1 The number sense: background

1.1.1 What is the "number sense"?

Sometimes at the state fair there is a game in which you guess how many candies are in a large jar, and the closest guess wins a prize. Of course, you cannot count the number of candies. You may use mathematical strategies to try to calculate a best guess, but most people rely on an underlying intuition. Just as we can take a stab at how much water will fill a barrel, how many square feet are in a room, or how loud a hand clap is, we can also estimate the number of items in a visual scene or the number of sounds in a sequence. Such estimates seem natural to us, and in some cases may seem more natural than physical quantities measured in units for which we have little intuition such as lumens or decibels. We refer to this intuitive approximate sense of numerical quantity as the "number sense" (Dehaene 1997).

It is tempting to view the number sense as rather mundane. Why should cognitive neuroscientists be interested in this type of intuition? In their influential review paper, Feigenson, Dehaene, and Spelke (2004) ask why mathematics is both so easy and transparent and yet also so difficult and academic. The proposition 1 + 1 = 2 can seem like a universally accessible *a priori* truth, and yet complex mathematics is so esoteric and specialized. They theorize that numerical concepts rely on two "core knowledge systems", one of which is the

approximate number intuition we are calling the number sense. Core knowledge systems are available from birth or develop very early in life and provide a set of intuitions about how the world works. Core knowledge systems can be combined and developed into more robust conceptual frameworks, especially using language. Although by itself the number sense intuition is not sufficient for a full concept of a “number”, by combining this intuition with other core knowledge systems and with language we can derive a conceptual system that is capable of sustaining symbolic arithmetic and mathematics. When math remains within the bounds of our core knowledge intuitions it is easy, but when it strays outside of them it becomes difficult. The number sense provides an intuitive bridge between the rarified realm of mathematics and the immersive world of sensory experience and may provide deeper lessons on how human concepts develop out of intuition.

1.1.2 Assessing the number sense without language: comparative and developmental perspectives

In the laboratory there are different ways we can assess the number sense. Perhaps the most straightforward is to ask people to guess how many items are in a visual array or auditory sequence (Izard and Dehaene 2008). For example, you can show an array of items on a computer screen and ask a participant, “How many items do you think there are?” Critically, to engage the number sense participants must be asked not to count, or the stimulus must be shown so briefly that counting is impossible. Unlike Dustin Hoffman’s character in

“Rain Man” and the poorly understood real savants with similar reported abilities (Sacks 1985), most people are pretty bad at this type of task. People’s estimates show a large degree of variability and generally underestimate the real value. Underestimation, however, seems to be merely a matter of “calibration”. When people are given a calibration array of 30 items and told explicitly that there are 30 items in that array, their subsequent estimates for a wide range of arrays improve dramatically, although never reach the level of savant ability (Izard and Dehaene 2008). This finding shows that an underlying mental representation of approximate number exists, but that its link or “calibration” to number symbols and words is tenuous and needs reinforcement.

We can study this underlying mental representation more systematically and directly by bypassing the need for verbal responses. Three perceptual tasks are common in the field. One is the ordinal comparison task, which we employ in all of the research chapters here. In ordinal comparison tasks, participants are instructed (or trained in the case of animals) to indicate either the ascending or descending numerical order of the stimuli. In most cases, this is simply a “choose greater” task in which participants must either pick the numerically greater of two stimulus arrays (as in the task in Chapter 3) or the more numerous color in an array of mixed items (Halberda, Mazzocco, and Feigenson 2008). Sometimes more complex responses are required such as indicating the correct ascending or descending order of multiple stimuli in a sequence (Brannon and Terrace 2000).

In addition to ordinal comparison tasks, another common type of numerosity task is match-to-numerical-sample. In this task, a sample array with a particular number of items is shown, followed by a test array. The participant must indicate whether the test array has the same number or a different number of items as the sample (Nieder, Freedman, and Miller 2002). A third type of non-verbal task for assessing the number sense is a production task, in which participants must generate a particular number of responses (J. R. Platt and Johnson 1971).

In general, these types of tasks allow us to more directly measure the underlying mental representation of magnitude without the added difficulty of estimating the effect of verbal or symbolic transformations. With the data from these tasks we can fit psychophysical models, the refinement of which is the goal of Chapter 3. Critically for the experiments here, these tasks also allow us to study the number sense in non-human animals lacking language.

Using the tasks outlined above, number sense abilities have been demonstrated in a variety of non-human animals. Chimpanzees are capable of ordinal comparison (Tomonaga 2007) and can even associate symbols with numbers (Matsuzawa 1985; Tomonaga and Matsuzawa 2002a). Macaque monkeys have been the most extensively studied; they can perform ordinal tasks (Brannon and Terrace 1998), delayed-match-to-numerical-sample (Nieder, Freedman, and Miller 2002), and production tasks (Sawamura, Shima, and Tanji 2002). Macaques are also able to make associations between numbers and

symbols (Livingstone, Srihasam, and Morocz 2010; Diester and Nieder 2007). Other primates, including new world monkeys (Judge, Evans, and Vyas 2005; Evans et al. 2009) and lemurs (Lewis, Jaffe, and Brannon 2005; Merritt et al. 2011), have succeeded at ordinal comparisons. Rats can perform ordinal tasks (Meck and Church 1983) and production tasks (J. R. Platt and Johnson 1971). Cetaceans can perform ordinal tasks (Kilian et al. 2003). There has been a less robust demonstration in cats (Pisa and Agrillo 2009). Outside the mammals, birds are also adept at ordinal tasks (Roberts 2010; Honig and Stewart 1989; Scarf, Hayne, and Colombo 2011) and matching tasks (Smirnova, Lazareva, and Zorina 2000), and even fish have shown proficiency on ordinal tasks (Agrillo, Dadda, and Bisazza 2006; Agrillo, Piffer, and Bisazza 2011). Thus, at least some form of the number sense appears to permeate the vertebrate taxa.

There is also evidence that when humans are prevented from counting, their performance on these tasks is very similar to animals. The most direct test of this hypothesis showed that, except for a small difference in acuity, monkey and human performance on an ordinal number task was very similar. In particular, as the numerical ratio of the arrays being compared approached one, performance dropped following a similar psychometric performance curve (Cantlon and Brannon 2006). This suggests that the number sense we experience is similar to the one experienced by animals.

The tasks used to measure the number sense in adults and animals must be modified for infants. Infants lack language like animals, but also lack the

motivation, motor coordination, and perhaps even the capacity for instrumental conditioning. As a result, implicit behavioral measures have been developed to gauge the number sense in infants. These approaches differ in their details, but in general take advantage of infants' interest novelty within the environment. The most common approach is a habituation dis-habituation paradigm (Xu and Spelke 2000). Infants are shown a series of visual arrays that differ in many features, but always have the same number of items. After some time, the infant loses interest in this stream of stimuli. At this point, the experimenter introduces new stimuli that differ in number. If infants observe a change in the stimulus stream, they will be more likely to refocus their attention to it. This allows experimenters to determine whether an infant can detect a particular difference across stimuli. Using this paradigm, researchers have found that adult humans and animals also share number sense with infants (Xu and Spelke 2000). Another paradigm used with infants consists of auditory stimuli in which the number of sounds is either matched or mismatched to the number of items in a visual array; using this paradigm, researchers found that newborns only a few hours old were able to detect differences in number (Izard et al. 2009). Unsurprisingly, infants' discrimination of different numbers is constrained and not as precise as it is in adults. In particular, the ratio of the numbers being compared influences discriminability. We can understand this intuitively rather easily. Although they both differ by one item, the difference between 2 and 3 items is much more obvious than the difference between 102 and 103 items.

During the first year of life, infants transition from being able to discriminate a 1:2 ratio to a 2:3 ratio (Lipton and Spelke 2003; Xu, Spelke, and Goddard 2005). Throughout childhood and into adulthood, this numerical ratio increases until it levels off in adulthood at around 8:9 (Halberda and Feigenson 2008a; Piazza et al. 2010).

Of course, children and adults show individual differences in their discriminable ratios. We can think of these differences as the acuity of the number sense. We will go into much greater detail regarding acuity below, but in general, people who have larger discrimination ratios possess a more acute approximate representation of number. Perhaps the strongest piece of evidence that the number sense provides a conceptual building block for mathematics is the recent finding that the acuity of the number sense in children is correlated with mathematical performance (Halberda, Mazzocco, and Feigenson 2008; Libertus, Feigenson, and Halberda 2013; Mazzocco, Feigenson, and Halberda 2011a). Children with a severe developmental delay in mathematics known as dyscalculia have a correspondingly less acute number sense than normal control children (Piazza et al. 2010). Discrimination acuity in preverbal infants has been found to influence mathematical proficiency in later childhood (Starr, Libertus, and Brannon 2013). This finding in particular suggests that the number sense serves as a conceptual building block in later math performance, because it rules out the possibility that previously demonstrated correlations were solely the result of mathematics learning improving number sense acuity.

1.1.3 The neurobiological basis of the number sense

Functional magnetic resonance imaging (fMRI) has shed light on the brain mechanisms of mathematical thought in general and the number sense in particular. Mental arithmetic produces a pattern of prefrontal and parietal activation (Dehaene et al. 1996). In particular, the intraparietal sulcus (IPS) is consistently activated during a variety of basic calculation and comparison tasks in adult humans (Pinel et al. 2001). Although many numerical tasks activate the IPS, some tasks activate it more and some less. Tasks involving the manipulation of quantities, a skill more closely associated with the number sense, is associated with greater activation of the IPS, whereas rote fact retrieval activates language areas like the left inferior prefrontal gyrus (Dehaene et al. 1999). Activity in the IPS is also modulated by the number of items in a visual array even when people are not engaged in any dot comparison task at all (Piazza et al. 2004; Harvey et al. 2013; Jacob and Nieder 2009). Interestingly, the IPS is activated when people view numerical symbols (Eger et al. 2003); when these symbols are varied, the same pattern of modulation is observed as when the number of dots in an array is modulated (Piazza et al. 2007).

Single cell recordings in macaque monkeys have furthered our understanding of how IPS activity is related to the number sense. Recordings from monkeys performing a delayed-match-to-numerical-sample task in the dorsolateral prefrontal cortex (DLPFC) and the IPS have revealed that the firing

rate of individual neurons is modulated by the number of items of visual dot array stimuli (Nieder, Freedman, and Miller 2002; Nieder and Miller 2004a).

The highest concentration of cells encoding numerical information recorded in the IPS were in the fundus of the sulcus (~20% of cell recorded), the ventral-most area of IPS known as the ventral intraparietal sulcus (Nieder and Miller 2004a). Subsequent studies have confirmed the presence of number neurons in VIP (Nieder 2012; Tudusciuc and Nieder 2007; Viswanathan and Nieder 2013).

The VIP neurons recorded during a numerical-match-to-sample task are tuned to individual numbers of items in the stimulus (Nieder and Miller 2004a). This means that the cell fires maximally for stimuli with a particular number of items and less for stimuli with fewer or more items. The further the number of items in a stimulus is from the preferred number, the less that cell will fire. Nieder and Miller observed cells tuned to numbers 1-5, the only values they tested. Later research found cells, at least in the DLPFC, that were tuned to a larger numerical range of up to 30 (Nieder and Merten 2007).

It should be noted that although most researchers interpret these data as supporting the theory that VIP neurons are tuned to individual numbers, there is another interpretation. It is possible that neurons are actually of just two types: those that increase their firing with number and those that decrease their firing with number. We will refer to these types of numerical representations as monotonic, to contrast it with the tuned representational model described above.

When the average firing rate is calculated for the presentation for each number of items, noisiness in firing rates can cause some monotonic neurons to appear as if an intermediate value triggered the maximal response. However, if enough trials were administered to pin-point the mean firing rates to each number precisely, it would be clear that all neurons respond maximally to either the smallest or largest number of items presented. This point is made in detail by Chen and Verguts (2013), and is discussed further in chapter 4.

A previous study from our laboratory recorded neurons in the lateral intraparietal area (LIP), the area of the IPS immediately lateral and dorsal to VIP, during the presentation of numerical stimuli (Roitman, Brannon, and Platt 2007). The number of items in the visual arrays was varied between 2 and 32, however no explicit numerical discrimination was required. The firing rates of over 50% of LIP neurons were found to be monotonically modulated (either up or down) by the number of items in the array.

The functional properties of macaque LIP and VIP have been well studied outside the context of the number sense. LIP is known to be involved in covert visual attention, saccade planning, and saccade initiation, with individual neurons tuned to particular locations within the visual field (Bisley and Goldberg 2003; Colby and Goldberg 1999; Colby, Duhamel, and Goldberg 1996). LIP also has an important role in value comparison and decision making (Platt and Glimcher 1999; Shadlen and Newsome 2001). It has been theorized that LIP is a salience map of the visual field, with visual attention focused wherever LIP activity is

greatest (Goldberg et al. 2006) or alternatively as encoding the subjective value of a particular choice target (Louie and Glimcher 2010).

VIP is characterized by strong motion direction tuning, including cells tuned to complex motion stimuli such as optic flow fields (Bremmer, Duhamel, et al. 2002). It is also a multimodal region with tactile, vestibular, auditory, and visual responses (Bremmer, Klam, et al. 2002; Duhamel, Colby, and Goldberg 1998; Schlack 2005). This pattern of response properties has been theorized to comprise a representation of self-motion during three-dimensional navigation (Bremmer, Klam, et al. 2002) or of approaching stimuli in peri-personal space (Graziano and Cooke 2006).

Some effort has been made to align the functionally defined macaque areas LIP and VIP to the areas identified in studies of numerical processing. Some recordings of number neurons in area VIP have found direction selective optic flow fields in the same locations where number neurons were recorded (Tudusciuc and Nieder 2007; Nieder, Diester, and Tudusciuc 2006). Roitman et al. (2007) used a delayed saccade task to ensure that the LIP number neurons recorded had a clearly defined spatial receptive field and pre-saccadic activity.

A review paper tried to tie multiple lines of human imaging evidence together (Hubbard et al. 2005). At the posterior occipital end of human IPS is putative human LIP activated by saccades. At the anterior end is the putative human anterior intraparietal area (AIP), associated with reaching and grasping. They found that the human IPS area most closely associated with calculation fell

along the human horizontal segment of the IPS. This area is much enlarged in humans, but generally falls between the grasping and saccade regions. The authors theorize that this area is most likely the human VIP, although the homology is tenuous.

1.1.4 Modeling the number sense

Meck and Church (1983) trained rats on a numerical bisection task using sequential auditory stimuli. The bisection task was similar to an ordinal task except only one sample stimulus was provided. During a training phase, rats were trained to press one lever when 2 tones were played and another lever when 8 tones were played. During the test phase, the training stimuli were intermixed with stimuli containing intermediate numbers of tones. They found that as the number of tones in a stimulus increased, the frequency of pressing the 8-tone lever instead of the 2-tone lever smoothly increased. Interestingly, at 4 tones, the rates were completely indifferent between the two levers, suggesting that the perceptual intermediate is the geometric mean.

To account for these findings, and others from related experiments on timing, Meck and Church proposed the mode control model. The model proposes a central pacemaker that produces “beats”. When the system detects an auditory tone, it “closes a switch”, allowing the pacemaker beats to enter an “accumulator”. The accumulator sums the new beats with any previously accumulated. At the end of the stimulus, a comparator compares the number of

beats in the accumulator with number of beats in memory for 8 tones and 2 tones. The rat then chooses the closest memory match for the current stimulus.

If the rate of the pacemaker were 100% reliable, this system would perfectly bisect stimuli around 4 tones, such that every number below 4 would be categorized as 2 and every number above 4 would be categorized as 8. However, data showed the frequency of picking 8 varied as a curvilinear function of the number of tones. To accommodate this, the authors posited that the pacemaker was “noisy”: on a given trial, the rate of the pacemaker was picked from a normal distribution. Although the underlying mean rate of the pacemaker was unknown, the ratio of the standard deviation to the mean rate could be fit to the data. This ratio, known as the coefficient of variation (COV), is a measure of the imprecision of the rat’s number sense. If the COV is high then the choice curve is shallow and more stimuli are difficult to classify. If the COV is low, then the choice curve is steep, and stimuli on either side of 4 are easily classified as more like 2 or more like 8. If the COV is zero, the classification is binary except for exactly 4, which is always classified with 50% probability.

Refinement of the accumulator model allowed it to be adapted to ordinal comparison tasks (Gallistel and Gelman 1992; Whalen, Gallistel, and Gelman 1999). These researchers demonstrated that when the COV is a constant ratio, which the authors term “scalar variability”, the accumulator model explains the “ratio effect”. The ratio effect is the main consequence of the Weber-Fechner law,

which states that the ratio of two physical magnitudes determines their discriminability.

The explanation is as follows. Gallistel and Gelman posit that the source of the noise in number sense representations is inherent to the representational format as opposed to being generated by an uneven pacemaker, as in the mode control model. These representations are themselves normal random variables. The standard deviation of a numerical representation is equal to its mean times the COV constant. As a result, larger numbers are represented by “fuzzier” representations with larger standard deviations. The likelihood of confusing two number sense representations is proportional to the overlap of their distributions. Interestingly, this overlap is constant for numbers separated by the same numerical ratio. Thus, the ratio effect in ordinal number sense tasks can be understood as a consequence of an internal numerical representation with a constant COV or scalar variability.

Further research demonstrated that when numerical stimuli are presented simultaneously as static arrays of items as opposed to sequentially presented tones or flashes, processing time does not depend on the absolute number of items being processed (Nieder and Miller 2004b). As a result, the models of the number sense, at least as applied to simultaneously presented stimuli, abandoned the concept of a pacemaker and accumulator inherited from the mode control model (Pica et al. 2004). This innovation was relatively minor since Gallistel and Gelman (1992) had already posited that the representation itself

was noisy, and all the same mathematical points regarding scalar variability hold. We are left to wonder, however, what the source of scalar noise is and whether some neurobiological explanation can be provided.

Piazza et al. (2004) provided further refinement to these psychophysical models. They noted that a different type of internal numerical representation could also explain the ratio effect. Instead of random variables with scalar variance, they posited log random variables with constant variance. When the distance between numbers is logarithmically compressed and all numbers are represented by normal distributions of constant variance, numbers separated by an equal ratio have an equal overlap, just as in the Gallistel and Gelman model. Figure 1-1 compares numerical representations posited by the Gallistel and Gelman model with the Piazza et al. model (from Cantlon et al. 2009).

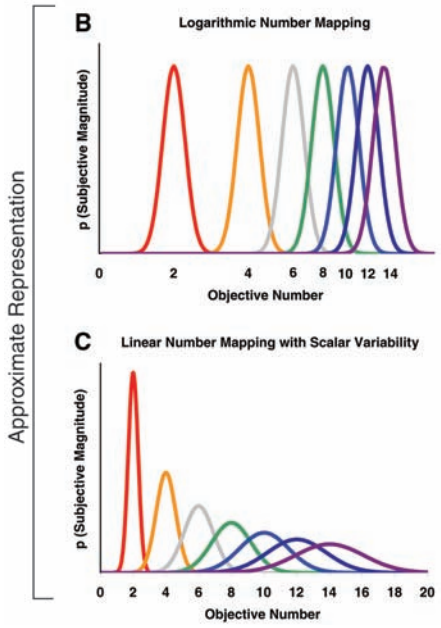

Figure 1-1. Hypothetical approximate mental representation of the even numbers 2-14 as posited by the log model with fixed variability and the linear model with scalar variability. (A) Excluded for brevity. Under the logarithmic code (B), numerical values are psychologically compressed logarithmically with a constant amount of noise. Under this system, numerical representations become increasingly less distinct as objective number increases because they become closer together in psychological space. The linear numerical code with scalar variability (C) represents numerical values with equal psychological distances between adjacent values, and the amount of noise in the numerical representation increases proportionally with its value. Like the logarithmic code, the linear-scalar code predicts that confusion between neighboring values increases with magnitude, not because of the subjective spacing of the values but because of the increased variability with which each value is represented. From (Cantlon et al. 2009). Reprinted with permission from AAAS and authors.

The predicted confusability of number representations in the logarithmic model with constant variance and in the linear model with scalar variance are almost identical (Cantlon et al. 2009). However, one theoretical advantage of the log model is it explains the indifference point observed by Meck and Church (1983). Recall that they found that rats reported the middle point between 2 and 8 as 4, the geometric mean, not 5, the arithmetic mean. They offered no explanation for this and simply built it into the comparator. If the mental representation of number is log compressed, the geometric mean is transformed

to the arithmetic mean, and the comparator can function by simply measuring Euclidean distance. For example, if we take the base 2 logarithm of 2, 4, and 8 we get 1, 2, and 3 respectively; obviously, 2 is the mid-point between 1 and 3, and no complex comparator characteristics are required.

The models outlined above posited an approximate internal numerical representation. Another line of modeling work has focused on creating artificial neural networks capable of extracting numerical representations from the visual scene. Dehaene and Changeux (1993) developed a four layer artificial neural network that extracted approximate number from a static visual array. The first layer consists of an artificial retina that simply registers the presence or absence of stimuli within the receptive field of each unit. The second layer performs normalization on object size, but preserves their location in the visual field. The third layer consists of a series of units all of which receive equal input from all second layer units, but each with an increasing activation threshold. As a result, when few items are presented, only the third layer units with the lowest thresholds are activated, but as more items are added, increasingly more third layer units are activated. The fourth layer consists of another series of units, each receiving input from one third layer input as well as inhibition from its neighbors. The result is that layer four units are tuned to individual numbers of dots.

For example, if 6 items are presented, there will be one unit in the third layer that has just barely reached threshold and become active. The corresponding fourth layer unit that receives input from this third layer unit will be

activated. It will be inhibited by third layer units with lower thresholds, but not by those with higher thresholds (because they themselves are not active). A fourth layer unit that receives activation from a third layer unit with a lower threshold will also receive inhibition from all of its neighbors and so will fail to activate. A fourth layer unit that receives input from a third layer unit with a higher threshold will receive no activation. Thus, for each number of items, there is a corresponding fourth layer unit. Further work demonstrated that this pattern of weights of layer three to layer four units can be generated readily using a simple learning algorithm (Verguts and Fias 2004).

The fourth layer of the neural network bears strong resemblance to the representation of number in the Gallistel and Gelman (1992) model. The activation of a unit linked with a particular number like 6 can be thought of as the representation of 6. If we consider that, instead of a clear binary between positive and negative weights in the third to fourth layer connection, there is a gradient, then we can imagine “fuzziness” in this representation. The “6 unit” would be partially activated by 5 and 7. Indeed, when tested, this neural network structure was capable of performing number discrimination tasks with response profiles like those observed in animals and humans. Critically, it reproduced the ratio dependent performance.

When Nieder et al. (2002) demonstrated neurons in DLPFC tuned to individual numbers of dots, the comparison to the Dehaene and Changeux (1993) model was apparent. These neurons and the similar ones found in IPS

are thought to comprise the instantiation of the fourth layer neurons in the model and form the neural basis of the number sense (Dehaene 2003). Similarly, the finding that LIP contains monotonic neurons seems to provide an analogue to the third layer neurons in the model, with the modification to include units that decrease activation with numerosity (Roitman, Brannon, and Platt 2007). Furthermore, findings of Nieder et al. (2002) were found to confirm the logarithmic model (Piazza et al. 2004): the response profiles of the neurons to different numbers of items were found to best conform to a normal distribution after logarithmic compression (Nieder and Miller 2003; Dehaene 2003). It should also be noted that it is not clear that tuned number neurons are not necessary for the performance of all number tasks; bisection tasks can theoretically be performed with just monotonic representations (Pearson et al. 2010).

1.1.5 Some terminology: "the approximate number system" and "numerosity"

Sometimes in the literature and throughout this dissertation the term "approximate number system" (ANS) is used. I do not mean to imply any particular model or theory with this term. Instead, I use it to describe perceptual processes that subserve the number sense, whatever they may be. The term is also not meant to imply that the processes subserving the number sense are fully modular and not related to or subsumed within other processes. Indeed, some of the data in the following chapters support the view that the number sense is closely related to other perceptual abilities.

Similarly, the word “numerosity” is sometimes used. It is used to refer to the internal representations of number that comprise the number sense. It is usually used to contrast a fuzzy approximate percept of numerosity with the fully defined symbolic number concept with which one can do mathematics. In this sense, numerosity is a building block for number.

1.1.6 Summary

The number sense is a core cognitive system. It is innate or develops very early in life. It is shared with many other animals and is likely evolutionarily ancient and genetically hardwired. In the IPS, numerosity is extracted from the visual scene by neurons tuned to individual numbers, which provide the basis for our sense of number. There is evidence that when number is relevant for behavior, the IPS forwards this information to other areas, such as the DLPFC, where numerosity can be used to guide behavior. This fundamental number sense may serve as a building block for more complex numerical concepts and mathematics.

1.2 Research questions

1.2.1 The question of malleability

It is not surprising given the summary view outlined above that little research has addressed the question of how experience shapes the number sense. If the number sense is genetically encoded and available from birth, then it is easy to dismiss a significant role for experience. All systems, however, develop, even those that have a strong genetic basis. Deprived of normal

experience, brain systems fail to develop normally. For example, if deprived of visual input from one eye, ocular dominance columns fail to develop in V1 (Wiesel and Hubel 1963). I do not think anyone would argue that visual perception in V1 and ocular dominance columns are therefore not innate. Furthermore, although the number sense emerges close to birth, its acuity continues to develop throughout infancy, childhood and adolescence (Piazza et al. 2010; Halberda and Feigenson 2008a). It is possible that some measure of plasticity still exists into adulthood.

Chapter 2 directly explores the question of adult number sense plasticity. We employed a training procedure with adult human participants to determine if number sense acuity could be improved. Chapter 4 indirectly explores the effects of experience on the neural representations in monkeys. Most previous studies of number neurons in VIP relied on well-trained animals (Nieder and Miller 2004a; Nieder 2012; Tudusciuc and Nieder 2007). In these experiments monkeys had tens of thousands of trials of experience on numerical tasks. These tasks explicitly reinforced attendance to the numerical dimension of stimuli. The experiment specifically targeting LIP did not use an explicit numerical discrimination task, but there was a predictable role of number (Roitman, Brannon, and Platt 2007). Stimuli of different numbers predicted differently sized rewards, and response time data demonstrated that monkeys did indeed attend to number. In the experiment in Chapter 4, we used monkeys that had never been trained on a numerical task. Thus, by comparing the types of

representations we found with those previously reported, we can make inferences regarding the effect of training on the neural code for numerosity.

1.2.2 The problem of stimulus control

The results from the experiment in Chapter 2 raised some important questions regarding the role of non-numerical stimulus features in numerical perception. In short, we found that the total area of the items in our stimulus arrays affected people's numerical intuition, and that training modulated this effect. At the same time we also found that training modulated numerical acuity. We wanted to disambiguate the two effects, which proved to be very difficult.

Controlling for the influence of non-numerical features in dot-array stimuli like those used in many studies of the number sense has long vexed researchers. If for example, you want to generate two arrays, one with two dots and one with eight dots, you have two choices about how to construct them. If you make the dots have the same average size in the two arrays, the total area taken by the dots will be four times greater in the eight-dot array. If, however, you fix the total area of the two arrays, the eight-dot array will have dots that are four times as small.

Most experimenters studying numerical cognition want to make sure that responses are based on number and not other visual features of the stimulus. To this end, they usually balance the two approaches outlined above. We were interested in measuring the influence of these other visual features, as well as

numerical acuity. Our efforts to differentiate these two effects led to significant modeling work and formed the basis of the study described in Chapter 3.

1.2.3 The question of IPS specialization

It is clear that the IPS is involved in cognitive processes besides the number sense. There is also strong evidence, especially in humans, of IPS specialization for number. The final question addressed in my dissertation is, to what degree is the IPS specialized for numerical processing in monkeys?

In Chapter 3 we gain powerful tools for disambiguating the effects of number and non-numerical features on the human number sense. In Chapter 4, we are able to apply those tools to the neural encoding of number in VIP. As a result, we are able to directly compare the encoding of number with the encoding of non-numerical features. We reasoned that if VIP were specialized for number, then number would influence firing rate more strongly and in more neurons than other visual features.

In Chapter 5 we test the question more directly. We use muscimol, a pharmacological agent, to reversibly inactivate LIP and VIP and observe the effect on the number sense and a color control task. We reasoned that if there is any specialization of IPS for number in monkeys, then we would see a greater decrement in performance on the number task than on control.

We will return to these questions in the conclusion to assess what answers we have and where future research might be directed.

2. Malleability of the approximate number system: effects of feedback and training

2.1 Introduction

Mathematics is a uniquely human domain because it requires symbolic manipulation and an explicit understanding of the operations that allow calculation. However, in addition to a symbolic number capacity, adult humans also have an approximate number sense that allows us to estimate quantity without the use of symbols or language. Unlike precise symbolic representations of individual numbers, the ANS encodes numerosities in a fuzzy fashion. A confluence of evidence suggests that the ANS emerges early in infancy and is shared by nonhuman animals (for reviews see Dehaene 1997; Feigenson, Dehaene, and Spelke 2004). One basic feature of the ANS is that it follows Weber's law; the discriminability of two numerosities varies as a function of the ratio between them.

During the course of normal human development the ANS becomes more precise. Convergent evidence from multiple behavioral procedures demonstrates that while 6-month-old human infants require a 1:2 ratio to discriminate large numerosities, by 9-months they are able to discriminate a 2:3 ratio (Libertus and Brannon 2010; Lipton and Spelke 2004). Furthermore, cross-sectional studies that model weber fraction (*w*) with explicit choice tasks indicate that the acuity of

the ANS continues to increase from age three into adolescence (Halberda and Feigenson 2008a; Piazza and Izard 2009). At each age, however, and into adulthood there exists a large amount of inter-individual variability in *w*. Not surprisingly, an easy numerical discrimination for one person may be difficult for another.

One dominant theory is that the ANS serves as a foundation for symbolic mathematics (Wynn 1998; Dehaene 1997). Psychophysical markers such as the symbolic distance effect indicate that mathematical symbols are mapped onto analog magnitudes (Moyer and Landauer 1967). Brain-imaging data indicate that symbolic calculations often activate the same brain areas involved in approximate estimation implying that the ANS is recruited during calculation (Venkatraman, Ansari, and Chee 2005; Fias et al. 2003; Holloway, Price, and Ansari 2010). Only recently, however, has evidence emerged that individual differences in *w* are correlated with symbolic mathematical abilities (Gilmore, McCarthy, and Spelke 2010; Halberda, Mazzocco, and Feigenson 2008; Lyons and Beilock 2011). These studies show that children and adults with higher ANS acuity (i.e., lower *w*) perform better in basic arithmetic and on standardized math tests. Furthermore, ANS acuity in preschoolers with no formal mathematics training correlates with later symbolic math performance implying that ANS acuity may play a causal role in the development of higher math skills (Mazzocco, Feigenson, and Halberda 2011b). Other evidence for the relationship between symbolic mathematics and number sense comes from atypically developing

children. Developmental dyscalculia is a specific learning deficit in mathematics, and there is evidence that some dyscalculic children have severely impaired ANS acuity (Mazzocco, Feigenson, and Halberda 2011a; Piazza et al. 2010). In addition, attempts to improve mathematical performance in dyscalculics that have centered on strengthening the connection between symbolic number representations and non-verbal numerosity representations (arrays of dots) have met with some success (Wilson, Revkin, et al. 2006; Kucian et al. 2011).

The fact that symbolic math ability and the ANS are correlated throughout childhood raises the exciting possibility that honing the ANS could have lasting effects on symbolic mathematics. If so, even before children learn the meaning of number words interventions that increase ANS acuity may produce increases in math aptitude. The idea that ANS acuity might serve a foundational role in developing mathematical achievement, however, cannot be addressed without a better characterization of the ANS. For example, how reliable are measures of ANS acuity and can ANS acuity be improved with extended training?

Another important question is how the ANS relates to the perception and discrimination of other magnitudes. Walsh (2003) proposed a theory of magnitude (ATOM), which asserts that time, space, and number are all processed by a common analog magnitude system that depends on common parietal brain systems (Cantlon, Platt, and Brannon 2009; Meck and Church 1983). A prediction of ATOM is that individual variability in the ANS should be systematically related to precision in other magnitude judgments (e.g., temporal

or size-based). A large literature addresses these questions in humans and animals using interference paradigms, transfer of learning tasks, and neuroimaging methods (for reviews see Hubbard et al. 2005; Bueti and Walsh 2009). Positive evidence from any of these sources could reflect a strong version of ATOM whereby two or more magnitudes are represented by a single common neural currency or a weaker version where different magnitudes share some common cognitive algorithms such as a comparison process (Cantlon, Platt, and Brannon 2009).

We explored the malleability of ANS acuity by testing whether a simple training procedure in which we provided extended training over six sessions would improve ANS acuity. We also tested a prediction of ATOM by looking for correlations in weber fractions derived from the numerosity comparison and those derived from a similar line length comparison task. In addition, our training paradigm allowed us to test a prediction of the strong version of ATOM by assessing whether improvements in the acuity of the ANS would transfer to line length comparison. We reasoned that if number and line length were represented using the same underlying representation (strong hypothesis), that any improvement in the number task would lead to an improvement in the line length task. If, however, we saw an improvement in ANS precision that did not transfer to the line length comparison we could conclude that the representations were not entirely overlapping, and more specifically the magnitude representations did not overlap on the level at which improvement occurred.

2.2 Materials and Methods

2.2.1 Participants

Participants were 20 adults (mean = 21.18 years, range: 18.19 - 30.15 years) recruited from the Duke University community. Eleven of the 20 participants were female. One additional participant was excluded because she did not receive feedback during the second session due to experimenter error. All participants gave written informed consent in accordance with a Duke IRB approved protocol.

2.2.2 Design

Each participant completed 6 sessions within 2 weeks. On session 1, participants performed the numerosity comparison task and the line length comparison task and did not receive any trial-by-trial feedback. On sessions 2-5 participants performed only the numerosity task and received trial-by-trial feedback. On session 6 participants performed the numerosity and line length tasks without trial-by-trial feedback. Each of the six sessions of the numerosity task contained six 108 trial blocks for a total of 648 trials per session. The two line length sessions each contained two 108 trial blocks for a total of 216 trials per session. In the final session participants self-reported their verbal and math SAT or GRE scores and these scores were later confirmed for 15/20 participants.

To motivate the participants to stay engaged in the task they were compensated based on performance. Each participant earned 0.0125 USD per correct answer in the numerosity task and 0.0375 USD per correct answer in the

line length task. These performance bonuses were added to a baseline rate of 7.50 USD for sessions 1 and 6 and 5 USD for sessions 2-5. To motivate participants to complete the study they were given an additional 50 USD for completing all 6 sessions within a 2-week period.

2.2.3 Tasks

The numerosity comparison task: On each trial participants were presented with an array of intermixed black and white dots on a gray background for 200 ms. Half the participants were instructed to indicate whether there were more black dots or white dots, and the other half were instructed to indicate whether there were fewer black dots or white dots. Participants responded by pressing a black or white button on the keyboard, and the side of the response keys was counterbalanced across subjects. Although the stimuli were presented for only 200 ms, participants were allowed as long as they needed to respond and were encouraged to take their time and to be as accurate as possible. In the feedback sessions, a green or red screen lasting 1500 ms indicated a correct or incorrect choice respectively. The feedback screen was followed by a grey preparatory screen (1500 ms). In the no-feedback sessions, any response resulted in a blue screen (1500 ms) followed by a grey preparatory screen (1500 ms). Participants were given a break between each 108 trial block and were also allowed to pause the experiment at any time.

The line length comparison task: The structure was similar to the numerosity task, however, subjects were presented with a white and a black

horizontal line and were required to indicate which was longer (or for half the subjects which was shorter) by pressing a black or white key. The line length comparison task was never administered with feedback, and it was only completed on sessions 1 and 6.

2.2.4 Stimuli and Apparatus

Each numerical stimulus consisted of an array of intermixed white and black dots (Figure 2-1). We tested six ratios of dots: 1:2, 2:3, 3:4, 5:6, 7:8, and 11:12. Absolute numerosity was roughly equated across the ratios, and the total number of dots within an array varied from 20 to 75. To ensure that subjects used numerosity and not surface area to complete the task on 1/3 of trials the total surface area of the array with fewer dots was smaller than the total surface area of the more numerous dots, on 1/3 of trials area was equal, and on 1/3 total surface area of the fewer dots was larger than the surface area of the more numerous dots. Similarly, to prevent subjects from using the size of the individual dots, on 1/3 of trials the average dot size of the fewer dots was smaller than the average dot size of the more numerous dots, and on 2/3 of trials the more numerous dots were smaller. The dots were drawn within a circle with a radius of 300 pixels.

The line stimuli consisted of one black and one white horizontal bar positioned at a constant vertical position (counterbalanced for which color was on top), but jittered horizontally from trial to trial (Figure 2-1). The same six ratios

were used for line lengths and numerosities. The length of the bars varied from 64 pixels to 384 pixels.

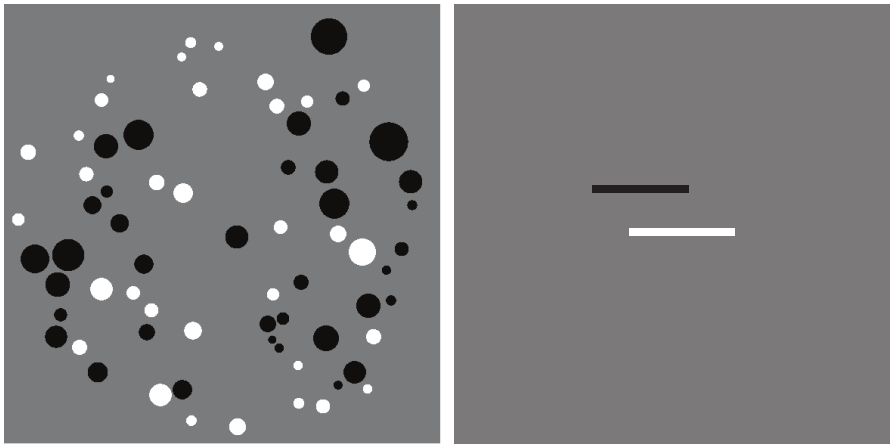

Figure 2-1. Example stimuli from the numerical (left) and line length (right) tasks
The numerical stimulus has 33 white dots and 36 black dots and is an example of an 11:12 ratio. The black line in the line length stimulus is smaller than the white line by a factor of 11:12.

All stimuli were generated offline using custom MATLAB (Mathworks) scripts, and were presented using Psychophysics Toolbox Version 3 for MATLAB. Stimuli were presented and data collected on either a Dell Inspiron 530S or a Dell Optiplex 330. Participants made their response on a standard keyboard. Small stickers were used to denote the 'black' and 'white' response keys.

2.2.5 Modeling

For the purpose of modeling we assumed a linear internal representation of number with scalar variability following Pica et al. (2004) and Halberda et al. (2008). The error rate in our task is given by

$$Error\ Rate = \frac{1}{2} \cdot erfc\left(\frac{n_1 - n_2}{w \cdot \sqrt{2} \cdot \sqrt{n_1^2 + n_2^2}}\right) \quad \textbf{(Eq 2-1)}$$

Where n_1 is the numerosity of the larger set, n_2 is the numerosity of the smaller set, *w* is the measure of variance in the internal representation, and erfc is the complementary error function. We generated global estimates of *w* for each participant as well as session by session estimates of *w* for each participant by fitting this model to our data (Pica et al., 2004). Figure 2-2 shows one participant's accuracy across different ratios and the model fit.

2.2.6 Surface area effect index

As described above we controlled for surface area by using three randomly intermixed trial types. To assess the role of surface area on performance we calculated a surface area effect index by taking the absolute value of the difference between the accuracy on the trials where the smaller number of dots had fewer pixels (congruent) and the accuracy on the trials where the smaller number of dots had more pixels (incongruent). We also calculated a non-rectified surface area effect index by computing the difference between congruent and incongruent trials, but not taking the absolute value. This measure allowed us to assess whether the population as a whole had a bias towards congruent or incongruent trials before and after training.

2.2.7 Treatment of outliers

Three of our participants (1, 5 and 16) returned *w* scores that were greater than 3 standard deviations above the mean (1.49, 1.23 and 1.36 respectively) for one of the six number sessions. On the other five sessions, these subjects' *w* values were within the same range as the other participants. We included these

participants in our main analyses, but also reran the statistics excluding these three subjects to confirm the robustness of our findings. Unless otherwise noted, all tests reported as significant were also significant without outliers at $p < 0.05$, and tests reported as non-significant were also not significant without outliers at $p > 0.1$.

2.3 Results

There was strong ratio dependence in accuracy ($b = -0.77$, $p << 0.0001$)[1] and response-time ($b = 0.53$, $p << 0.0001$) for the numerosity comparison task. The mean *w* for the sample was 0.33 with a standard deviation of 0.15. Within session reliability was computed by correlating split-halves of our six blocks and correcting for test length using the Spearman-Brown formula. Reliability was good, ranging from 0.83 to 0.94 over the six sessions. Our multi-session training procedure allowed us to further examine test-retest reliability across the 6 days of the study. Figure 2-3 shows the strong positive correlation between *w* scores computed from the first session and last session.

[1] There was no difference in *w* scores for subjects instructed to indicate the greater versus the fewer number of dots (t-test, $p = 0.62$) thus all analyses are collapsed across these two groups.

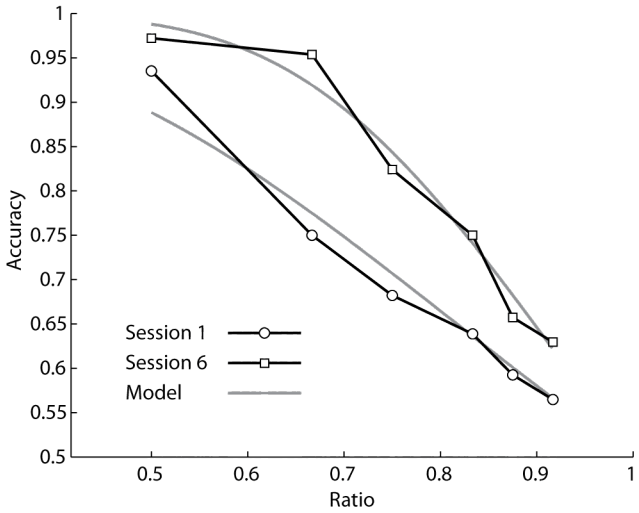

Figure 2-2. Raw data and model fit for a single participant.
Data collected from a single participant showing improvement in number acuity between session 1 (circles) and session 6 (squares). Gray lines show model predictions for the best fit *w* for session 1 (*w* = 0.37) and session 6 (*w* = 0.20). The improvement in *w* was typical of our sample.

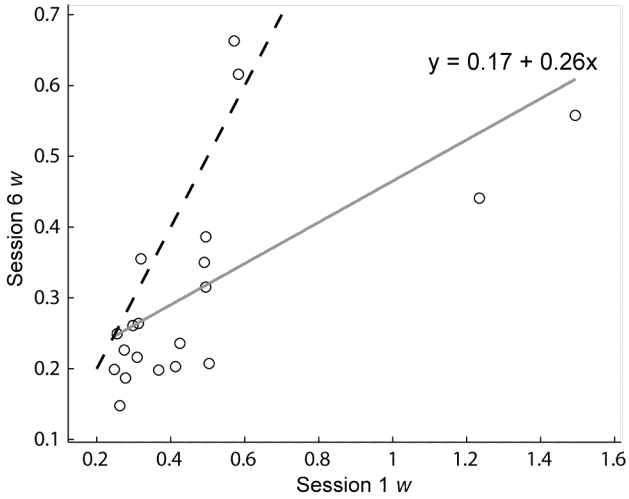

Figure 2-3. Numerosity *w* scores calculated from session 6 plotted against numerosity *w* scores calculated from session 1.
The dashed line shows equality. Participants below the equality line showed improvement in ANS acuity between sessions 1 and 6. The solid grey line is the best fit to the data and the equation shows the intercept and slope (r^2 = 0.40, p < 0.005).

To determine if *w* scores improved (decreased) with training we calculated *w* scores for each participant for each session. We then fit a logarithmic regression model to individual *w* scores with regressors for session number and participant (Figure 2-4A). The model accurately predicted *w* scores (R^2 = 0.72, p

<< 0.0001), and we found that *w* scores improved with training (b = -0.047, p < 0.005). However, the improvement in *w* scores occurred within the second session (the first session with feedback) and remained stable during the rest of training. Session 1 *w* scores were significantly higher than session 2 *w* scores (paired t-test, p < 0.005) and were also higher than session 6 *w* scores (paired t-test, p < 0.01). A logarithmic regression model applied to the session 2 through session 6 data showed no effect of session on *w* (b = 0.008, p = 0.65) demonstrating that the improvement in *w* was accomplished within the first session of trial-by-trial feedback and did not continue with extended training. Figure 2-2 shows the accuracy data and model fit of a single participant for sessions 1 and 6. The improvement in *w* was typical of our sample.

To determine if response time decreased with training we fit a logarithmic regression model with regressors for session number and participant to the median correct RT calculated for each subject for each session (Figure 2-4B). The model accurately predicted RT (R^2 = 0.83, p << 0.0001), and RT decreased with training (b = -0.048, p << 0.0001). Unlike *w*, however, RT continued to decrease from session 2 to 6 (b = -0.035, p < 0.0001). RT rebounded slightly on session 6 when feedback was removed. Nevertheless RT during session 6 was significantly lower than on session 1 indicating that the improvement was retained in the absence of feedback (paired t-test, p < 0.005).

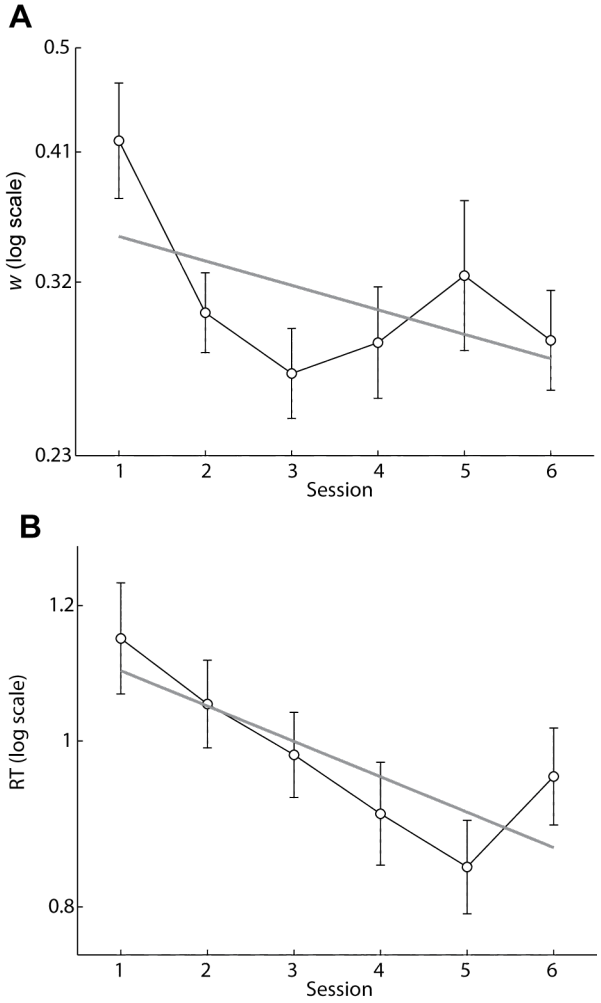

Figure 2-4. Mean *w* and RT scores.
(A) Mean *w* scores as a function of session number. Grey line is the best fit regression line to log(*w*) (b = -0.047, p < 0.005) accounting for repeated measures. Note the rapid decrease in *w* between sessions 1 and 2 where trial-by-trial feedback was introduced. Acuity did not improve further after session 2, but was sustained after feedback was removed on session 6. (B) Mean of the median RT for correct responses as a function of session number. Grey line is the best fit regression line to log(RT) (b = -0.048, p << 0.0001) accounting for repeated measures. RT continued to decrease with further training. Error bars indicate SEMs.

To measure the effect of cumulative surface area on participants' numerical estimation over training a surface area effect index was calculated by taking the absolute value of the difference between the accuracy on trials where the smaller number of dots had fewer pixels (congruent trials) and the accuracy on trials where the smaller number of dots had more pixels (incongruent trials). A

linear regression accounting for repeated measures was then fit to the surface area index (R^2 = 0.49, p << .0001). The surface area effect index significantly decreased over sessions (b = -0.02, p < 0.005; Figure 2-5), indicating that at least part of the improvement in the numerosity task was due to a decrease in reliance on surface area as a cue for number. There was, however, variability across participants in the degree to which surface area affected numerosity judgments and also in the direction of this influence. During the first session most participants performed better on area congruent than on area incongruent trials. The mean non-rectified surface area effect index was significantly positive (mean = 0.24, std = 0.27; t-test, p < 0.001), indicating higher accuracy on congruent trials. However, by the last session participants performed equally well on both types of trials (mean = -0.07, std = 0.22; t-test, p = 0.17). Surprisingly, when outliers were removed from this analysis the non-rectified surface area index was slightly negative indicating that subjects performed better on incongruent trials by the last session (mean = -0.09, std = 0.16; t-test, p < 0.05).

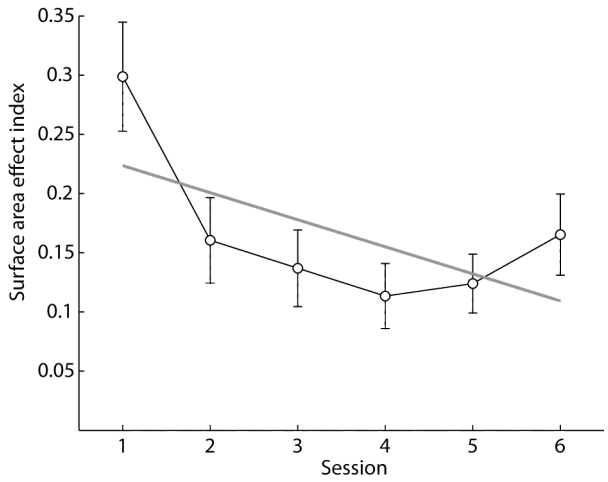

Figure 2-5. Mean surface area effect index as a function of session number.
The effect of surface area on accuracy decreased rapidly with the introduction of trial-by-trial feedback, plateaued after session 2, but remained low after feedback was removed on session 6. Grey bar indicates the linear best fit accounting for repeated measures. Error bars indicate SEMs.

On a third of trials the cumulative surface area of the white dots and the black dots was equal. We looked at accuracy on this subset of trials in order to ascertain whether the improvement in *w* we observed was due solely to the decrease in bias caused by surface area, or whether other factors might also be contributing to improvement. We found that accuracy on area equal trials was well fit by a linear regression ($R^2 = 0.67$; $p << 0.0001$) and trended towards a significant positive slope ($b = 0.0043$; $p < 0.1$; without outliers $p < 0.05$). Closer examination of the data showed that the effect was not linear over sessions, but that all the improvement occurred between sessions 1 and 2. We ran a 2-way ANOVA with factors for session and participant to confirm the effect of session on equal area accuracy ($F(5,95) = 4.81$; $p < 0.001$). In post-hoc t-tests we found that accuracy on area equal trials increased between sessions 1 and 6 (paired t-

test, $p < 0.005$) and between sessions 1 and 2 (paired t-test, $p < 0.001$), but not between sessions 2 and 6 (paired t-test, $p = 0.80$). The rapid increase in accuracy between sessions 1 and 2 demonstrates a comparable time course to our findings for *w* and for the surface area effect index.

Overall, line length *w* scores (mean = 0.07, std = 0.02) were much lower than number *w* scores (mean = 0.33, std = 0.15) indicating that the line length task was easier. Despite having different absolute ranges, *w* for line length and numerosity were positively correlated on session 1 (Figure 2-6A, $r^2 = 0.44$, $p < 0.005$) and on session 6 ($r^2 = 0.38$, $p < 0.005$) consistent with a weak version of ATOM. Line length comparison tests were only given on the first and last session and subjects were never given trial-by-trial feedback on this task. Thus, any improvement from session 1 to session 6 on the line length task could be attributed to training on the number task and would thus reflect transfer across magnitudes as predicted by a strong version of ATOM. However, a comparison of *w* scores from the first session and the last session yielded no evidence of improvement in line length acuity (one-tailed paired t-test: $p = 0.283$). We examined the relationship between change in acuity on the numerosity task and change in acuity on the line length task in individual participants, but found no correlation (Figure 2-6B, $r^2 = 0.00$, $p = 0.99$) indicating that subjects who improved on the number task were no more or less likely to have improved on the line length task.

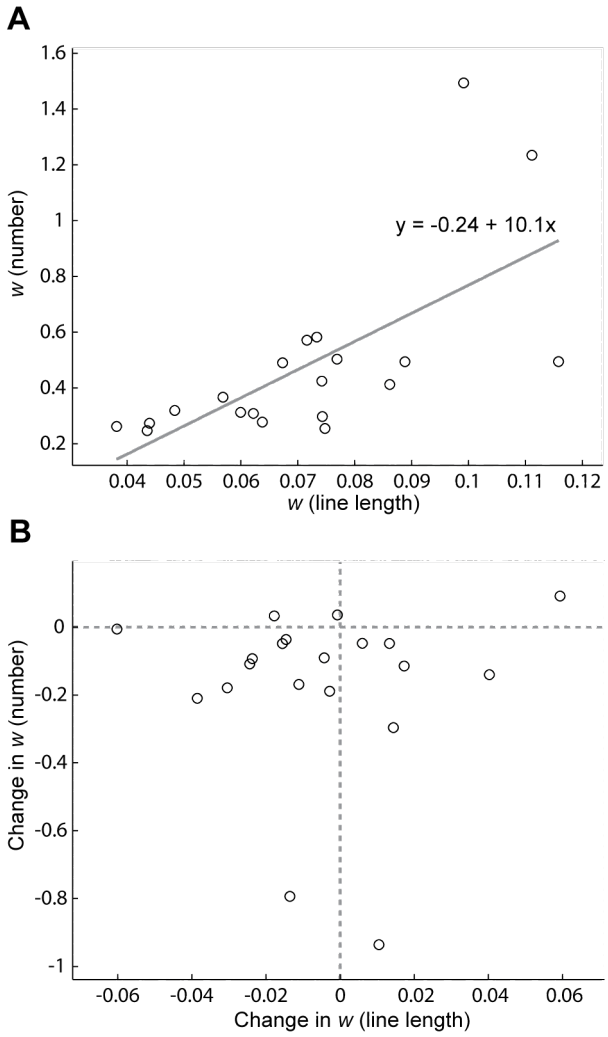

Figure 2-6.
(A) Numerosity *w* scores were positively correlated with line length *w* scores on session 1 (r^2 = 0.44, $p < 0.005$) and session 6 (r^2 = 0.38, $p < 0.005$, data not shown). Grey line is the best fit to the data and the equation shows the intercept and slope. (B) Change from session 1 to session 6 in numerosity *w* scores plotted against the change in line length *w* scores over the same period. Numerosity *w* scores were not correlated with any improvement in line length *w* scores (r^2 = 0.00, p = 0.99). Note that most participants (17/20) had a negative change in ANS *w* indicating an improvement. Change in line length *w*, however, was evenly distributed around zero indicating no improvement in line length acuity in the population. Data points to the left of the vertical dotted line indicate an increase in line length comparison acuity whereas data points to the right indicate a decrease in line length comparison acuity from session 1 to session 6. Data points below the horizontal dotted line indicate an increase in numerosity comparison acuity whereas data points above indicate a decrease in numerosity comparison acuity from session 1 to session 6.

Accuracy on the line length task was very high on both the first and last sessions (mean correct = 93.6% and 93.9% respectively), which may have

created a ceiling effect that obscured any improvement on the line length task from sessions 1 to 6. We addressed this concern by assessing change in accuracy on only the most difficult 11:12 ratio line length comparison (mean accuracy 80.1% std = 10.6% and 84.6% std = 10.4% during sessions 1 and 6 respectively). Consistent with the original analysis, we found no evidence of improvement in accuracy on this subset of trials between sessions 1 and 6 (one-tailed paired t-test: $p = 0.12$). This confirms that number training caused no detectable improvement in line length acuity. Furthermore, there was no correlation between improvement in *w* scores for the number task and change in accuracy on these most difficult line length comparisons ($r^2 = 0.01$, $p = 0.68$).

Previous reports have demonstrated that standardized math scores correlate with numerical acuity in children (Halberda, Mazzocco, and Feigenson 2008; Gilmore, McCarthy, and Spelke 2010; Mazzocco, Feigenson, and Halberda 2011b). One recent study also showed a positive correlation between *w* and mental arithmetic in adults however the relationship was mediated by ordinal symbol knowledge (Lyons and Beilock 2011). Consistent with these reports we found a negative correlation between SAT/ GRE score and *w* (Figure 2-7A, $r^2 = 0.28$, $p < 0.02$) and no correlation between verbal SAT/ GRE score and *w* (Figure 2-7B $r^2 = 0.08$, $p = 0.23$). This negative correlation did not hold when the three participants with single-session outlier data were excluded (without outliers: $r^2 = 0.04$, $p = 0.47$). However, when *w* was recalculated for these three subjects excluding the single session for which each subject exhibited an outlier *w* score

the negative correlation was significant with math SAT/ GRE scores (r^2 = 0.27, p < 0.05), but not verbal scores (r^2 = 0.08, p = 0.22).

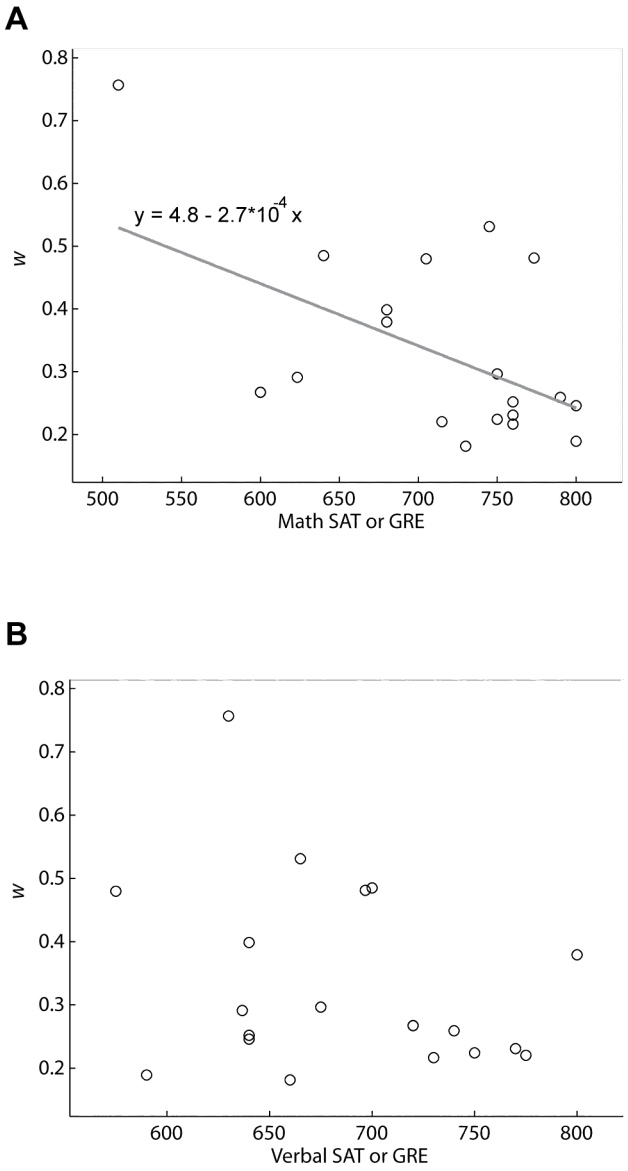

Figure 2-7. Correlation between w and standardized math and verbal tests.
(A) Standardized mathematics test scores (GRE or SAT) were negatively correlated with *w* (r^2 = 0.28, p < 0.02). Grey line is the best fit to the data and the equation shows the intercept and slope. (B) Verbal scores were not significantly correlated with *w* (r^2 = 0.08 p = 0.23).

2.4 Discussion

2.4.1 Malleability of ANS acuity

The primary question our research addressed was the malleability of the weber fraction in response to extended training. We found rapid improvement in ANS acuity with the introduction of trial-by-trial feedback and this improved performance was maintained in a final session when feedback was omitted. Very little improvement in ANS acuity occurred after the first session in which trial-by-trial feedback was introduced (second actual session) suggesting that ANS acuity may plateau and then be insensitive to extended training. Response time, however, continued to decrease with further training on the task. It remains possible that the four sessions of training with feedback that we provided was not sufficient and that additional training would have reduced the weber fraction further even in these participants. It is also possible that extended training with feedback in children who have not yet reached asymptotic performance in ANS acuity would be more effective and we plan to pursue this in future research.

Why was the introduction of trial-by-trial feedback so powerful in reducing the weber fraction? One caveat is that our study did not include a control group that did not get feedback. Therefore it is possible that initial practice, and not trial-by-trial feedback was the main factor in reducing *w* in the first session of the number task. Future studies should explore this possibility. Another possibility is that feedback allowed subjects to decrease reliance on total stimulus surface area. The effect of surface area and numerosity congruency was strong in the

majority of subjects before trial-by-trial feedback was introduced. Thus subjects tended to view arrays with larger total surface area as more numerous. With training, however, the effect of surface area decreased, and by the final session participants no longer showed a surface area bias. The decrease in the congruence effect, however, cannot fully explain the observed decrease in *w*. On trials were the surface area of the two arrays was equal, we still observed an increase in accuracy after feedback was introduced, and, like the effect seen in *w*, this improvement in accuracy persisted after feedback was removed.

Decreasing reliance on total surface area as a mechanism for improving ANS acuity is consistent with theories of perceptual learning. Goldstone (1998) identified attentional weighting and differentiation as potential mechanisms for perceptual learning. Changes in attentional weighting can allow participants to focus on crucial information like numerosity while ignoring irrelevant stimulus features like surface area. Differentiation allows previously indistinguishable aspects of stimuli to be perceived as distinct and has been shown to apply to different perceptual dimensions of the same stimulus. For example, according to the Munsell color system colors vary along three orthogonal dimensions: chroma, value, and hue. (Burns and Shepp 1988) found that trained subjects were significantly better at differentiating value and chroma than untrained subjects. Similarly, subjects trained to categorize color based on chroma but not value increased their acuity in discriminating different chroma (Goldstone 1994). These results suggest that our participants may be learning to differentiate the related

dimensions of numerosity and surface area allowing them to ignore the extraneous surface area cues and to selectively improve number acuity.

Prior studies have examined the relationship between surface area and perceived numerosity in adults and come to different conclusions. Consistent with our findings, Hurewitz, Gelman, and Schnitzer (2006), found that congruence between surface area and number improved accuracy whereas incongruence caused a decrement in performance. Others, however, found the opposite effect, that larger items were perceived as less numerous (Tokita and Ishiguchi 2010). Barth (2008) failed to find any effect of surface area congruence in an ordinal numerosity task. As Tokita and Ishiguchi (2010) demonstrated and we confirm here, trial-by-trial feedback rapidly diminishes or abolishes surface area bias. It remains an open question, however, exactly what stimulus or presentation factors determine the direction or existence of surface area bias effects in naïve subjects. One potentially important difference between our study and the Tokita and Ishiguchi study was that we presented dot arrays simultaneously and spatially overlapped whereas they used sequential presentation.

It is interesting to note that studies with children suggest that the ability to separate dimensions improves with age (Smith and Evans 1989; for review see Goldstone 1998). Thus children may be more susceptible to the surface area numerosity congruence effect than adults, and this effect may diminish with development and increasing acuity of the ANS. A large literature addresses the

effect of surface area on number judgments across development. However, there is no consensus on how these interactions change with experience (e.g. Piaget 1965; Mix, Huttenlocher, and Levine 2002; Cantlon, Safford, and Brannon 2010).

2.4.2 The ANS and other magnitude systems

A second question our study addressed was the relationship between ANS acuity and the precision of line length comparisons. We found that performance on a line length task was positively correlated with performance on the ANS task. The introduction of feedback on the numerosity task, however, improved acuity for the numerosity task but did not generalize to the line length discrimination.

Walsh's theory of magnitude (ATOM) asserts that dimensions such as time, number, and space are processed by a common analog magnitude system and depend on a common set of parietal brain systems (Bueti and Walsh 2009; Walsh 2003). The association between the spatial and numerical dimensions has been particularly well established (for review see Hubbard et al. 2005). Many studies have demonstrated interference between numerical and spatial information, the SNARC effect being the most well-known (Dehaene, Bossini, and Giraux 1993). Parietal lesions causing hemi-spatial neglect often cause congruent neglect in the mental number line, implicating common parietal circuits in both spatial and numerical cognition (Cappelletti et al. 2007; Zorzi, Priftis, and Umiltà 2002). Disruption of normal parietal function with rTMS causes deficits in comparing line lengths and numerosities (Dormal, Andres, and Pesenti 2012a).

Brain imaging studies have also implicated overlapping areas of the parietal cortex in both length and numerical comparison tasks (Fias et al. 2003; Dormal and Pesenti 2009).

Our finding that line length acuity correlated with ANS is consistent with the theory that spatial and numerical comparisons depend on shared cognitive mechanisms. However, the improvement that emerged from the introduction of trial-by-trial feedback did not transfer to the line length task. This finding is consistent with a weaker version of ATOM in which magnitude comparisons share some common basis but at least in adulthood are differentiated. One possible explanation of this partial differentiation is that a single common comparator system is utilized in all judgments of relative magnitude regardless of dimension, but that each magnitude is represented by a dimension specific subsystem. Thus, although number and line length are represented along distinct mental continua, comparisons of two numbers or line lengths are mediated by a single common comparator. Under this framework, the correlation between ANS acuity and line length acuity is explained by the resolution of a common comparator. In contrast, trial-by-trial feedback in the numerosity comparison task results in improvements that are specific to numerosity representations (e.g., increasing precision of the underlying representations or narrowing in of attention to the numerosity dimension as opposed to surface area). Future work might be able to disentangle the effects specific to a mental magnitude comparison and the precision of representations of a specific mental magnitude by, for example,

comparing the accuracy of a numerosity estimation task (how many dots?) and a numerosity comparison task, like the one we used. If the underlying representation of number narrows due to training, then it should to transfer across different number tasks.

An important caveat is that our control task had some significant limitations. One limitation was that we were only able to test one non-numerical magnitude judgment (i.e., line-length), and we did not assess a non-magnitude perceptual judgment. This prevented us from determining whether the correlation between the number and line length weber fractions was due to global cognitive influences such as attention or fatigue, or alternatively arose from common magnitude processing mechanisms. Furthermore, we equated the ratios for the numerical and line length stimuli and this meant that the line length stimuli were significantly easier to discriminate than the numerical stimuli. One reason for this apparent disparity in difficulty may be that to solve the numerical task participants had to ignore total surface area, which was carefully controlled, whereas in the line length task there was no competing dimension. However, when we analyzed the most difficult line length trials we found to improvement in accuracy indicating that the lack of transfer was not due to a ceiling effect in the line length task. Future studies should include additional control tasks and match difficulty and stimulus complexity to make firmer conclusions about the import of the positive correlation we observed between ANS and line length judgments.

2.4.3 Relationship between the ANS and symbolic math

A third question our findings address is the relationship between ANS acuity and symbolic mathematics. Recent work has demonstrated that ANS acuity is positively correlated with a variety of mathematical abilities in children and adults (Halberda, Mazzocco, and Feigenson 2008; Gilmore, McCarthy, and Spelke 2010; Lyons and Beilock 2011; Mazzocco, Feigenson, and Halberda 2011b; Mazzocco, Feigenson, and Halberda 2011a). These studies suggest the ANS may serve as a developmental building block upon which symbols are mapped and that precision in ANS representations facilitates symbolic mathematics (e.g. Verguts and Fias 2008; Dehaene 1997; Mundy and Gilmore 2009; Gilmore, McCarthy, and Spelke 2007; Wynn 1998). A great deal of work is still needed to probe the dynamics of this relationship and to specify the mechanisms by which ANS acuity might scaffold symbolic mathematics. Consistent with these prior recent studies, our sample of adult participants exhibited a positive correlation between ANS acuity and standardized math scores but not verbal scores. Future work should explore the functional relationship between the ANS and mathematics by assessing whether improving ANS acuity, perhaps earlier in development, bestows any benefits for symbolic mathematics (e.g. Wilson, Revkin, et al. 2006; Wilson, Dehaene, et al. 2006; Kucian et al. 2011).

There are several possible explanations for why we did not find a more robust relationship between standardized mathematics scores and *w*. We had to

combine SAT scores with GRE scores, since a few of our participants had not taken the SAT. Although the tests are similar and graded on the same scale (200 – 800 points), combining GREs and SATs certainly added noise to the measure. In addition, our sample did not contain much variance in math scores, and may have suffered from a ceiling effect. Only one participant had a math score below 600, whereas fully half our sample scored 750 or above. Thus future studies should recruit larger samples from a more heterogeneous population.

2.4.4 Absolute value and reliability of w

Global *w* scores for our sample fell between 0.18 and 0.76 with a mean of 0.33 and a standard deviation of 0.15. This is higher than most previous estimates for young adults, which cluster below 0.2 (for review and meta-analysis see Piazza and Izard 2009) but was similar to the range of 0.22 to 1.5 measured by Gilmore et al. (Gilmore, Attridge, and Inglis 2011) in their non-symbolic comparison task. The disparate ranges in these three studies are surprising given the similarity of the estimation tasks.

One possible reason we observed higher *w* is that we did not control the dot density of our stimuli. Our two stimuli were generated within a single circle 300 pixels in radius. As a result the total extent of each stimulus was equal, but the density of the stimulus was negatively correlated with numerosity. Previous research has demonstrated that loosely spaced dots appear greater in number than densely packed dots (Ginsburg 1976). If the density of each set of dots was viewed independently (e.g. adding black dots did not increase the perceived

density of the white dots) then this effect may have inflated estimates of our less numerous stimuli, which would have appeared less densely packed and therefore more numerous thus impairing discriminability. Lower accuracy would have increased our estimate of *w*. This effect may have been especially pronounced in our stimuli because they had a relatively large degree of visual crowding. Further research into the specific effects relative density and other low level stimulus features on ANS acuity may help clarify differences in average *w* in different experiments.

We also measured the reliability of *w* scores by comparing split halves of individual session data. Single session reliability estimates were high and similar to estimates obtained in previous reports (Gilmore, Attridge, and Inglis 2011; Maloney et al. 2010). Note that one other report obtained low estimates of split-half reliability, however they used the distance effect rather than *w* as a measure of ANS acuity (Sasanguie et al. 2011). Our repeated testing design allowed us to assess reliability in *w* across six sessions over a 2-week period. Despite the reduction in *w* from session 1 to session 6 there was strong positive correlation in these scores demonstrating test-retest reliability over a 2-week period. These data thus provide evidence of both stable and malleable components of ANS acuity.

2.4.5 Potential single neuron correlates

Single cells in the monkey brain appear to encode quantity. The firing rate of cells in or near the intraparietal sulcus in macaque monkeys are systematically

correlated with the numerosity of dot arrays (Nieder and Miller 2004a; Roitman, Brannon, and Platt 2007), the numerosity of sequential actions (Sawamura, Shima, and Tanji 2002) or sequentially presented stimuli (Nieder, Diester, and Tudusciuc 2006), and line length (Tudusciuc and Nieder 2007). Prefrontal cortical cells also encode numerosity (Nieder, Freedman, and Miller 2002) and more abstract magnitudes such as symbol numerosity mappings (Diester and Nieder 2007), and ordinal rules (Bongard and Nieder 2010).

There are several different ways in which we can imagine single cell number coding systems to yield improved performance as a result of training. Neurons found in the IPS and PFC are tuned to individual numerosities. Tuned number neurons fire maximally for a particular numerosity and decrease firing in response more distant numerosities. One possibility is that the behavioral improvements we observed as a result of trial-by-trial feedback are achieved by sharpening the tuning curves of these neurons. As a result they would fire less for neighboring numerosities and be more selective for their preferred numerosity after training. Alternatively, training and feedback may recruit more individual neurons to the representation of number. This could improve the precision of the population code without affecting the width of the tuning curves of individual number selective neurons.

Other neurons in lateral intraparietal area (LIP) have been shown to encode numerosity monotonically, with separate populations either increasing or decreasing firing rate with the observed numerosity (Roitman, Brannon, and Platt

2007). Monotonic numerosity neurons have been hypothesized to play the role of numerosity accumulators in several models of numerical cognition (Dehaene and Changeux 1993; Meck and Church 1983; Verguts and Fias 2004). The accumulation layer in these models plays an intermediary role between perception of the stimulus and the final tuned representations of individual numerosities. Improved performance as a result of training might emerge from a sharpening of these accumulator-like neurons in LIP. After training, a given difference in numerosity would generate a greater increase (or decrease) in the firing rate in LIP neurons. An increase in the steepness of these monotonic functions could increase discriminability between numerosities and in turn lead to sharper tuning functions in downstream areas, including other areas in the IPS and in prefrontal cortex. Pearson et al. (2010) demonstrated that LIP like monotonic functions are in principle sufficient for completing a numerosity bisection task. This raises the possibility that different numerosity representations may be generated idiosyncratically in response to particular task demands, and training and education may play an important role in determining which types of number representations become realized in the brain.

Tudusciuc and Nieder (2007) found both line length and numerosity neurons in macaque intraparietal cortex. However, they did not find neurons representing magnitude abstractly along a common mental magnitude line. Line length and numerosity were represented in separate neuronal populations. A small percentage of neurons represented both line length and numerosity, but

these neurons were tuned to different line length and numerosity magnitudes: a neuron that coded for a short line length was equally likely to code for a small or a large numerosity. Thus, in monkeys it seems that line length and numerosity magnitudes do not share a common encoding scheme on the single neuron level in IPS. If one of the mechanisms of acuity improvement outlined above selectively acted on the numerosity neurons in the IPS but not the line length neurons, this could explain the failure of acuity improvement to transfer from one magnitude dimension to another.

2.4.6 Conclusion

Our study addressed the malleability of the ANS and the relationship between the ANS and other judgments. We found that ANS acuity showed rapid improvement with the introduction of trial-by-trial feedback but that it was otherwise relatively impervious to extended training in adults. The improvement in *w* in response to feedback was at least partially due to a decrease in reliance on surface area as a cue for numerosity, although other factors also influenced improvement. Acuity in a line length discrimination was positively correlated with ANS acuity, however, improvement in the ANS in response to feedback did not transfer to improvement in this spatial magnitude discrimination, providing further evidence that magnitude judgments may have both shared and distinct components. Finally, even in our relatively small sample of 20 subjects, acuity of the ANS was positively correlated with standardized tests of mathematical but not verbal proficiency. These findings raise important questions about the

malleability of the ANS over the lifespan and the relationship between the ANS and uniquely human mathematical abilities.

3. Modeling the approximate number system to quantify the contribution of visual stimulus features

3.1 Introduction

The approximate number system (ANS) is a nonverbal mechanism for estimating the number of items in a set, that develops early in human ontogeny and is shared with a wide array of animals (Feigenson, Dehaene, and Spelke 2004). The ANS is faster but much less accurate than verbal counting. The ANS may serve as a neural scaffold for symbolic mathematics, a proposition supported by the finding that ANS acuity (*w*) predicts math achievement in both children and adults (DeWind and Brannon 2012; Gilmore et al. 2013; Gilmore, McCarthy, and Spelke 2010; Halberda et al. 2012; Halberda, Mazzocco, and Feigenson 2008; Lyons and Beilock 2011; Mazzocco, Feigenson, and Halberda 2011a; Piazza et al. 2010; Starr, Libertus, and Brannon 2013; but see Holloway and Ansari 2009; Sasanguie, Defever, et al. 2013; Sasanguie, Göbel, et al. 2013) and that extensive practice on tasks that tap the ANS improves symbolic math performance (Hyde, Khanum, and Spelke 2014; Park and Brannon 2013).

The acuity of the ANS typically is measured by presenting arrays of dots and requiring participants to indicate which has more. When dot arrays differ in numerosity, however, other properties of the stimuli—such as dot size, dot

density, and array extent—differ as well. Many prior studies have found that non-numerical visual stimulus features influence numerosity discrimination performance, thus interfering with precise estimates of ANS acuity (DeWind and Brannon 2012; Frith and Frith 1972; Gebuis and Gevers 2011; Ginsburg 1976; Sophian 2007; Tokita and Ishiguchi 2010).

While most researchers acknowledge that non-numerical stimulus features influence numerosity judgments, the two most commonly used models of ANS acuity do not account for these biases (Piazza et al. 2004; Pica et al. 2004; Whalen, Gallistel, and Gelman 1999). These models postulate that numerosity is represented by the ANS as a normally distributed random variable with a mean equal to the number being represented and a width proportional to ANS acuity (*w*). Errors in numerical discrimination occur when the numbers of items being compared activate overlapping internal numerosity representations. According to these models the overlap of these representations is entirely attributable to the ratio of the numerosities being compared and the *w* term.

When the *w* parameter in these models is fit to accuracy data from a dot array comparison task all errors in the task are implicitly assumed to be the result of imprecision of the representation of number. However, since non-numerical features also affect numerosity judgments they sometimes cause errors (or correct responses) that cannot be attributed to numerical ratio. These responses are incorrectly attributed to imprecision or precision in the representation of number. As a result, the *w* measure derived from the current models of numerical

representation conflates the acuity of the numerical representation with the biasing effects that non-numerical stimulus features have on that representation. In practice, this means *w* is influenced by idiosyncrasies in the way the experimenter has chosen to control for non-numerical stimulus features. In its most extreme form large differences in the congruence or incongruence of non-numerical stimulus features with number results in wildly divergent estimates of *w* in the same individual, causing some to question the existence of the ANS independent of non-numerical feature cues (Szucs, Nobes, et al. 2013).

Here we introduce a new "stimulus space" that elucidates the dependencies and degrees of freedom inherent in dot array stimuli. Utilizing the insights provided by the stimulus space, we then propose a modification to the logarithmic model of the ANS that explicitly accounts for the effects of non-numerical stimulus features on numerosity judgments. This approach allows ANS acuity to be estimated independently of the influence of multiple non-numerical stimulus features, thus yielding a more reliable and more theoretically valid estimate of *w*. Improved estimation of *w* will help researchers elucidate relative importance of ANS acuity on mathematical cognition, the factors that may mediate that relationship, and its developmental trajectory.

In addition to making theoretical advances in modeling *w*, our model also returns coefficients describing the influence of non-numerical stimulus features that provide novel quantitative parameters useful for comparing individuals. We assessed the prevalence of non-numerical feature bias among educated adults

and statistically tested the hypothesis implicit in the current models of the ANS: that *w* and numerical ratio are the only factors that determine the discriminability of dot arrays in a numerical discrimination task. The non-numerical feature coefficients also provide a straightforward and quantitative way to assess the use of “alternative strategies”, that is the reliance on non-numerical features instead of numerosity to make discriminations between stimuli. Here we are able to provide a comprehensive unbiased assessment of the role of ten non-numerical stimulus features in numerical discriminations and to test the hypothesis that the ability to approximately enumerate is reducible to co-varying non-numerical cues.

Here we model choice behavior in adults based on the number, size, and spacing of dots, however, with slight modification we can use those same factors to model neural dependent variables such as neuronal firing rate, electroencephalography (EEG) scalp voltage, or blood-oxygen-level-dependent (BOLD) signal. Thus, in addition to clarifying the effect of different stimulus features on behavior, our new modeling approach can help elucidate which brain responses reflect number as opposed to other features of a stimulus.

3.2 Theory and calculations

We applied a novel analytical technique to model numerical discrimination performance as a function of numerosity, item size, and item spacing. Our approach relies on the insight that although arrays of dots have many different features that all covary, the features known to influence numerosity judgments have three degrees of freedom. Thus, numerosity discrimination performance

can be modeled as a function of just three stimulus features: the number of dots in the array and two novel parameters that describe the size and spacing of the dots within the array. From the coefficients returned for these three features the influence of many other non-numerical features can then be calculated. This modeling approach allows a dissociation of ANS acuity from the biasing effects of non-numerical visual features, thus yielding a theoretically valid estimate of ANS acuity that is more reliable across different stimulus sets.

3.2.1 Intrinsic and extrinsic stimulus features

Our approach requires a full understanding of the relationship among numerosity, intrinsic features of the stimulus, and extrinsic features of the stimulus (Dehaene, Izard, and Piazza 2005; Piazza et al. 2004). Intrinsic features are parameters of the individual items within an array, whereas extrinsic features are parameters of the array as a whole. When the numerosity of an array is fixed, the relationship between a given pair of intrinsic and extrinsic features is linear. For example, total surface area (an extrinsic feature) is equal to the number of items multiplied by the item surface area (an intrinsic feature). The same relationship exists between field area (the space within which the dots are drawn, sometimes referred to as the envelope or the convex hull) and sparsity (average field area per item, or the inverse of the density). For a given numerosity, increasing sparsity necessitates a linear increase in field area. Another way of describing these relations is to say that numerosity, item surface area, total surface area, sparsity, and field area are not mutually independent of each other,

and describing all of them overdetermines the stimulus. A smaller subset of these features are sufficient to determine the full set of features, an idea we will return to below.

Figures 3-1A and 3-1B plot “stimulus spaces” that summarize these relationships. Stimulus parameters are plotted with intrinsic features on the x-axis and extrinsic features on the y-axis. Figure 3-1A shows the intrinsic and extrinsic features related to the size of the items, item surface area and total surface area, and Figure 3-1B shows the intrinsic and extrinsic features related to the spacing of the items, sparsity and field area. Also apparent are what we term iso-numerosity lines (gray); all stimuli of a particular numerosity lie on a single iso-numerosity line, the slope of which is equal to the numerosity. Different points along an iso-numerosity line correspond to stimuli that differ in the intrinsic and extrinsic properties but have the same numerosity. An individual stimulus occupies a single point in both Figure 3-1A and Figure 3-1B, for example, the stimulus labeled S1 has a numerosity of 8. Location of a single stimulus in each of the two plots is constrained by numerosity; it must fall along the same iso-numerosity line in both plots. However, its location along that line is independent in each of the plots. In other words, the size of the items and spacing of the items in an array are independent of each other.

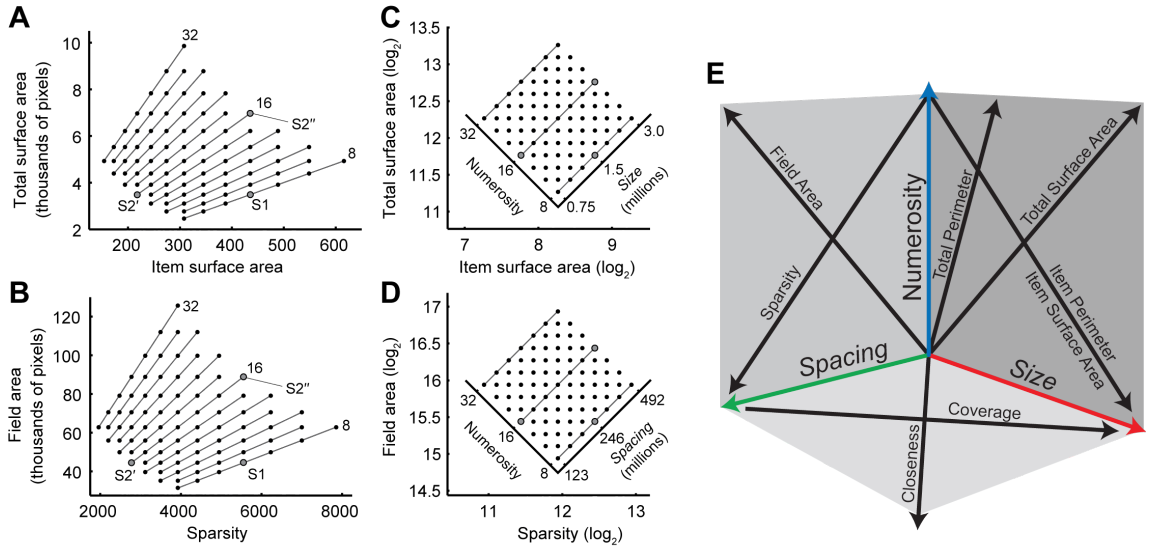

Figure 3-1. Features thought to influence numerical estimation can represented as different axes in a three dimensional stimulus space.
(A and B) Stimuli (black dots) used in this experiment plotted as item surface area by total surface area (A) and sparsity by field area (B). Stimuli of the same numerosity fall along iso-numeral lines (gray). The slope of these lines is equal to the numerosity of the stimuli that fall along them. Example stimuli referred to in the text are labeled and indicated by gray dots with black outlines. (C and D) The same plots as in (A) and (B) but with x and y axes log scaled. Changes in numerosity occur along a linear axis and two orthogonal dimension, *Size* and *Spacing* are apparent. The alternative internal axis dimensions provide an equally descriptive quantitative account of stimulus features as the external axes do. (E) The log of numerosity, *Size*, and *Spacing* plotted as cardinal axes of a 3D stimulus space. Log of non-numerical stimulus features are also plotted as arrows to indicate the direction in the space in which they increase. Any dot array stimulus can be uniquely defined with respect to numerosity, *Size*, *Spacing*, item surface area, total surface area, item perimeter, total perimeter, field area, sparsity (and density), coverage, and closeness on the basis of its location in this space.

The essential confound inherent in generating pairs of numerosity stimuli can be appreciated visually in Figures 3-1A and 3-1B. Two stimuli must be chosen, each from a different iso-numerosity line, however, any two such stimuli will also differ in the intrinsic variable, the extrinsic variable or both. It is mathematically impossible for two stimuli to differ only in numerosity. For example, consider again the stimulus labeled S1 in Figures 3-1A and 3-1B of numerosity 8. We may want to pair this stimulus with another of numerosity 16 in an ordinal comparison task, and seek a way to control for changes in other visual

features (the intrinsic and extrinsic variables). Now consider the two stimuli labeled S2′ and S2″, both of numerosity 16. S2′ has the same total surface area and field area as S1, but a different item surface area and sparsity. In contrast, S2″ has the same item surface area and sparsity as S1, but a different total surface area and field area. Other stimuli occupying other positions along the 16 iso-numerosity line would differ in both intrinsic and extrinsic stimulus features. All stimuli that differ in numerosity from S1 must also differ in either item surface area or total surface area and must also differ in either sparsity or field area.

3.2.2 Logarithmic scaling and deriving orthogonal regressors

Figures 3-1A and 3-1B are redrawn in Figures 3-1C and 3-1D, but with the intrinsic and extrinsic axes log scaled. Log scaling the axis affords two critical advantages: it makes the iso-numerosity lines parallel (for clarity only the iso-numerosity lines for 8, 16, and 32 are shown), and it makes the distance between stimulus points in the space proportional to the ratios of their various features (numerical and non-numerical). As a result, changes in numerosity are represented as movement along a linear numerosity dimension, and changes in the extrinsic or intrinsic features that do not result in a change in numerosity are represented as movement along an orthogonal linear axis. We represent these linear orthogonal stimulus dimensions as alternative axes in Figures 3-1C and 3-1D to emphasize that they represent a quantitative way to specify the location of a stimulus in “stimulus space”, which contains all the same information as specifying the values of the intrinsic and extrinsic features. Furthermore, these

three alternative axes represent a minimally sufficient set of features for describing the numerosity, as well as both sets of intrinsic and extrinsic features of a stimulus. For the rest of this manuscript we will refer to these linear orthogonal dimensions as *Size* for the intrinsic and extrinsic variables item surface area and total surface area and *Spacing* for sparsity and field area. We use capitalization and italics to make it clear that we are referring to a rigorously defined mathematical construct, but we also wish to emphasize the close relationship of these terms to the everyday concepts of size and spacing. Intuitively, changes in *Size* are equivalent to changing the size of a fixed number of items with fixed distances between their centers, and changes in *Spacing* are equivalent to changing the distances between a fixed number of items of fixed size.

We algebraically define *Size* and *Spacing* by examining the relationship between intrinsic and extrinsic stimulus parameters, and finding the dimension that is orthogonal to numerosity.

$$log_2(n) = log_2\left(\frac{TSA}{ISA}\right) = log_2\left(\frac{FA}{Spar}\right) \quad \text{(Eq 3-1)}$$

Where *n* is the number of items, *TSA* is the total surface area, *ISA* is the item surface area, *FA* is the field area, and *Spar* is the sparsity. The dimensions orthogonal to log number are

$$log_2(Size) = log_2(TSA) + log_2(ISA) \quad \text{(Eq 3-2)}$$

$$log_2(Spacing) = log_2(FA) + log_2(Spar) \quad \text{(Eq 3-3)}$$

Size and *Spacing* capture the aspects of dot size and spacing that are independent of numerosity. These definitions support the basic logic of the regression model that we will formulate in section 3.3.6. The effect of the numerical ratio on choice behavior can be assessed as in previous models, by fitting the *w* term. We can expand that model, however, by adding terms that will quantify the effect of the *Size* ratio and *Spacing* ratio. By including terms, which capture the perceptual effects of item size and spacing that are independent of the perceptual effect of numerosity itself, we lay the basis for independently assessing their contributions to numerosity discrimination performance.

3.2.3 Non-numerical stimulus features can be reduced to linear combinations of numerosity, Size, and Spacing in a logarithmic stimulus space

Figure 3-1E represents numerosity, *Size*, and *Spacing* as cardinal axes in a three-dimensional stimulus space with log scaled axes. We can imagine generating this space by taking the two two-dimensional spaces in Figures 3-1C and 3-1D and intersecting them along the numerosity axis at right angles to each other (in the third dimension). Any given array of dots is described by a single position within this three-dimensional space. The three dimensions (log of numerosity, log of *Size*, and log of *Spacing*) are independent of each other, and these three variables fully determine the extrinsic and intrinsic parameters discussed above (as illustrated in Figures 3-1C and 3-1D). Importantly, several other stimulus features are also fully specified by numerosity, *Size*, and *Spacing*. For example, the item perimeter and total perimeter are determined by the

numerosity and *Size* parameters. Coverage and apparent closeness are two features that depend on *Size* and *Spacing* and are not related to the numerosity of the stimulus. Coverage, sometimes referred to as density (Gebuis and Reynvoet 2011), is the total surface area per field area. Apparent closeness is the overall scaling of the stimulus, and increasing it is equivalent to zooming in on a stimulus such that it subtends a larger visual angle without changing its relative proportions. Appendix A contains the equations that relate each of these features to numerosity, *Size*, and *Spacing*. These equations demonstrate that our stimulus space is very descriptive. With just three numbers it specifies a stimulus's numerosity, item surface area, total surface area, sparsity (and density), field area, item perimeter, total perimeter, coverage, apparent closeness, *Size*, and *Spacing*. This descriptiveness is important for the modeling approach that we describe below, because it provides the basis by which our model can infer the effect of any stimulus feature on discrimination performance while only containing terms for numerosity, *Size*, and *Spacing*.

It is worth noting that log scaling is not merely a mathematical trick; it has important neurobiological and theoretical bases as well. Response function of neurons tuned to individual numerosities found in prefrontal cortex in monkeys are logarithmically compressed (Nieder and Miller 2003). Theoretically, the Weber-Fechner law states that the discriminability of two stimuli is linearly related to their ratio, equivalent to their distance on the logarithmic scale. For example, according to Weber-Fechner, a stimulus of numerosity 8 and one of numerosity

16 are equally discriminable as a stimulus of 16 and one of 32, because both pairs have a 1:2 ratio. In logarithmic stimulus space the distances along the numerosity dimension between 8 and 16 and between 16 and 32 are equal. Thus, we can use the difference in log numerosity as a regressor in a generalized linear model of numerical discrimination. Indeed, this is the approach of the logarithmic model developed by Piazza et al. (2004).

On a logarithmic scale, the equations relating the numerosity, *Size*, and *Spacing* to the other non-numerical features have a critical feature: they are all linear equations (Appendix A). Geometrically, this means that the dimensions along which different non-numerical features increase are straight lines in the three-dimensional stimulus space illustrated in Figure 3-1E. Furthermore, the distance along any of these dimensions that separates a pair of stimuli is proportional to the ratio difference of that feature. Thus, one of the benefits of the new stimulus space introduced here is that just as a 1:2 ratio of numerosity corresponds to a fixed distance along the numerosity dimension, here a fixed ratio of any feature corresponds to a fixed distance along its own dimension. For example, the distance between two stimuli along the total surface area dimension, one of which is comprised of 2000 pixels and the other of 4000 pixels, will be the same as the distance between a stimulus comprised of 4000 pixels and one of 8000 pixels, since both these stimulus pairs differ by a 1:2 total surface area ratio. Therefore, we can extend the logic of the Piazza et al. (2004) model to non-numerical stimulus features. Instead of assuming that only the log

of the numerical ratio affects numerical judgments, we can determine which stimulus feature ratios are affecting judgments, and we can do so in a manner that does not favor any particular feature.

It may be tempting to simply include regressors for the log ratio of all of the non-numerical features in a generalized linear model and have them compete with numerosity to explain the variance in behavioral discrimination performance. Although any two stimulus features are only partially collinear, some combinations of two features are fully collinear with a third, making such a model overspecified. Instead, we can take advantage of the linear equations that relate the log ratios of all the other features to the log ratios of numerosity, *Size*, and *Spacing* (Appendix A). These linear relationships mean that we can use the log of the S1 to S2 numerosity ratio, the log of the S1 to S2 *Size* ratio, and the log of the S1 to S2 *Spacing* ratio as regressors in a generalized linear model of numerosity discrimination. From the coefficients returned we can then infer the effect of a ratio difference of any feature on numerosity discrimination performance.

In short, previous models of numerical comparison predict that accuracy in a numerical ordinal comparison task is a function of the numerical ratio. The model, which we introduce in the methods section below, allows instead that accuracy is a function of the numerical ratio, the *Size* ratio, and the *Spacing* ratio. The effects of *Size* and *Spacing* ratio on accuracy would be of little interest by themselves, since they are merely novel mathematical constructs. However, by

virtue of the relationship between *Size* and *Spacing* and the other non-numerical features, estimating the effect of *Size* and *Spacing* on accuracy is mathematically equivalent to estimating the effects of all the non-numerical stimulus features on accuracy.

To evaluate our new modeling approach, we tested 20 adult participants using a standard non-symbolic numerical ordinal comparison task. Participants were instructed to choose the array that contained more dots. Our stimuli were generated such that numerical ratio, *Size* ratio, and *Spacing* ratio were varied independently across stimulus pairs. We developed a generalized linear model that allowed us to fit choice curves that modeled each participants' sensitivity to each of those ratios. We hypothesized that numerosity ratio would be the main determinant of choices given the task instructions, but that *Size* and *Spacing* ratio would influence choices.

3.3. Methods and Materials

3.3.1 Participants

Participants were 20 adults (mean 22.9 years, range 19.6 – 26.8 years) recruited from the Duke University community. Eleven of the 20 participants were female. All participants gave written informed consent in accordance with a Duke IRB approved protocol.

3.3.2 Design

Five participants completed ten sessions within 11 days and performed a maximum of three sessions in one day. Another 15 participants completed a

single session in one day. Each session lasted about 1-hour and consisted of 750 trials broken into three blocks of 250 trials each. Participants were required to take a five-minute break between blocks. Participants were compensated 10 USD for each session.

3.3.3 Task

Participants were seated in front of a computer and instructed to indicate the side of the screen that contained the greater number of dots using the arrow keys on a standard keyboard. Instructions were given verbally at the beginning of the session and in written format on the computer screen at the beginning of each block. At the beginning of each trial, a readiness cue was presented in the center of the screen (500 ms) followed by two arrays of white dots on a black background presented simultaneously to the right and left of the readiness cue (eccentricity ~8.5 degrees) for 250 ms. A response prompt was then presented followed by a 2 second inter-trial interval.

Participants were given eight easy practice trials (1:4 numerical ratio) at the beginning of each block. Practice trials were identical to the experimental trials except that they had a longer readiness cue time (1 s), longer stimulus display time (1 s) and a longer inter-trial interval (4 s). In the rare event that a participant responded incorrectly on any practice trial the script terminated with a prompt to see the experimenter. The experimenter then repeated the instructions and the block restarted.

3.3.4 Stimuli

We constructed a stimulus set that divided two octaves of numerosity, *Size*, and *Spacing* into 13 levels, evenly spaced on a logarithmic scale. The range of stimulus parameters is shown in Figure 3-1A through 3-1D, with all of the stimuli plotted in the stimulus space described in section 2.2. For 7 of the 13 numerosities, stimuli were generated at 7 different *Sizes* and 7 different *Spacings* yielding a total of 7 x 7 x 7 = 343 stimuli. For the other 6 numerosities, stimuli were generated at 6 *Sizes* and 6 *Spacings* for a total of 6 x 6 x 6 = 216 stimuli. Thus there were 559 unique stimulus parameter combinations. On each trial the experimental program randomly picked one of 4 different numerical ratios (closest whole numbers to $1:2^{1/6}$,$1:2^{1/3}$,$1:2^{1/2}$, or 1:2 ratios), one of 13 *Size* ratios (all possible pairings), and one of 13 *Spacing* ratios (all possible pairings).

In order to spread stimuli evenly along a logarithmic scale the values were rounded to the nearest whole number. For example, 32 is 2^5 and 8 is 2^3. However, we wanted 11 more powers of two spaced evenly between 2^3 and 2^5, such as $2^{4.5}$. $2^{4.5}$ is approximately 22.627 which we rounded to 23. Similarly, dot diameters and field diameters were rounded to the nearest whole pixel so they could be drawn properly on a monitor.

After defining the number, *Size*, and *Spacing* of a stimulus the algorithm created an instantiation of that stimulus. First the field area and item surface area were calculated (see Appendix A for the relations between number Size and Spacing and the other visual magnitudes). Dots of the appropriate size were

drawn at random locations within a circular field of the appropriate area. The only constraint on placement was that all dots were separated by at least one pixel and that all the dots were completely within the circle defining the stimulus field. It is worth noting that the circular field was not necessarily the smallest circle that could encompass all the dots in the array, although across multiple stimuli the field area and the smallest encompassing circle area were closely correlated.

3.3.5 Modeling choice behavior with existing models

We compared our model to the two standard models for estimating numerical acuity (*w*). The first model, termed here the "logarithmic model," assumes numerosities are represented as normally distributed random variables on a log compressed mental number line with means equal to the logarithm of the number represented and a fixed standard deviation (Piazza et al. 2010; Piazza et al. 2004). That model was used to fit data in which participants' compared a deviant value to a fixed standard value (either 16 or 32). The probability of choosing "larger" for the deviant stimulus was the proportion of the numerosity distribution lying on the greater side of the standard. The probability of this occurring at different numerosities is a cumulative normal distribution with a standard deviation that is equal to the standard deviation of the representation of numerosity, *w*.

In contrast participants in our task were asked to pick the larger of two numerosities that both varied from trial to trial with no fixed reference value. To accommodate this change in paradigm, we modified the model used by Piazza et

al. (2010) to include two numerosity distributions on a log-compressed number line, each with equal variance *w*. According to this version of the logarithmic model, the probability of choosing a stimulus was the proportion of its numerosity distribution lying on the greater side of the other stimulus distribution (not a fixed referent). The probability of this happening at different log right to left numerosity ratios is a cumulative normal distribution with standard deviation of *w* multiplied by root two.

$$p(ChooseRight) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{log_2(r_{num})}{\sqrt{2}(\sqrt{2}w)}\right)\right] \quad \text{(Eq 3-4)}$$

Where r_{num} is the ratio of the right side to the left side stimulus and *erf* is the error function.

The second model, termed here the "linear model," assumes that number is represented on a mental number line that is linearly spaced but has variance that scales linearly with magnitude (Halberda, Mazzocco, and Feigenson 2008; Pica et al. 2004). In this model *w* is the scalar that relates the numerosity to the standard deviation.

$$p(ChooseRight) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{(r_{num}-1)}{\sqrt{2}w\sqrt{{r_{num}}^2+1}}\right)\right] \quad \text{(Eq 3-5)}$$

It should be noted that *w* refers to different mathematical constructs in the logarithmic and linear models, making direct comparisons meaningless. Indeed, the same accuracy data fit by these two models produces different numerical values for *w*.

3.3.6 A novel model of numerosity discrimination that accounts for the effect of non-numerical features

We compared the two models above with the model we developed that was designed to accommodate the empirical fact that the size and spacing of dots within an array affect subjective numerosity. We fit a generalized linear model to choice data with regressors for the log of the ratio of numerosity, *Size*, and *Spacing* of the stimulus appearing on the right and the stimulus appearing on the left. The model formatted as a linear equation:

$$p(ChooseRight) =$$

$$(1-\gamma)\left(\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{\beta_{side}+\beta_{num}log_2(r_{num})+\beta_{Size}log_2(r_{Size})+\beta_{Spacing}log_2(r_{Spacing})}{\sqrt{2}}\right)\right]-\frac{1}{2}\right)+\frac{1}{2}$$

(Eq 3-6)

This equation looks rather different, but it can be thought of as simply an elaboration of equation **(4)**. This can be better appreciated if we rearrange it:

$$p(ChooseRight) =$$

$$(1-\gamma)\left(\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{log_2(r_{num})-\left(\frac{-\beta_{side}-\beta_{Size}log_2(r_{Size})-\beta_{Spacing}log_2(r_{Spacing})}{\beta_{num}}\right)}{\sqrt{2}\frac{1}{\beta_{num}}}\right)\right]-\frac{1}{2}\right)+\frac{1}{2}$$

(Eq 3-7)

The $\log_2(r_{num})$ term is equivalent between equation **(4)** and equation **(7)**. The standard deviation of equation **(4)** was *w* multiplied by root two, whereas here the standard deviation is the reciprocal of β_{num}:

$$\sigma = \frac{1}{\beta_{num}} \quad \text{(Eq 3-8)}$$

Therefore we can compute *w* from the new model parameters according to:

$$w = \frac{\sigma}{\sqrt{2}} = \frac{1}{\sqrt{2}\beta_{num}} \quad \text{(Eq 3-9)}$$

This makes sense: the *w* term depends only on β_{num}, the term capturing participants' sensitivity to number. There are also meaningful differences between the models. In equation **(4)** $\log_2(r_{num})$ alone determines the x-axis position along a single choice curve where the indifference point is at 0 (a 1:1 ratio). The greater the numerical ratio of right to left the greater the probability of choosing right. In equation **(7)** this is still true, but there is now a large term subtracted from the x-axis position. This value determines indifference point of the choice curve, which can now vary according to several new terms. In the context of a cumulative normal choice curve the indifference point is the mean and it is given by:

$$\mu = \frac{-\beta_{side} - \beta_{Size} log_2(r_{Size}) - \beta_{Spacing} log_2(r_{Spacing})}{\beta_{num}} \quad \text{(Eq 3-10)}$$

These new terms include the log ratios of *Size* and *Spacing* as well as all the β terms. β_{side} is an offset term that accounts for any side bias a participant might have. β_{Size} and $\beta_{Spacing}$ modulate the degree to which the *Size* and *Spacing* ratios affect the indifference point, and β_{num} scales the effect of all factors such that the greater the numerical acuity the smaller the effect of everything else.

The other new term in the model is the γ. Because our task was fast paced to allow many trials to be collected within a reasonable amount of time, we assumed that participants occasionally looked away from the screen or were momentarily distracted and failed to process the stimuli. In this case participants'

choices would be random and not related to any stimulus characteristics. To accommodate this we included γ a guessing term (Halberda and Feigenson 2008b; Pica et al. 2004). This term allows choice curves to asymptote below 100% and above 0%, since the more a participant guesses the more the entire choice curve is compressed towards 50%. In the extreme example of a participant who responded randomly the γ term would be 1, the proportion of rightward responses would be 50% and no other term in the model would matter.

It is worth noting that if γ and all the β terms besides β_{num} are zero the new model completely reduces to the logarithmic model in equation **(4)**. However, if the β terms for *Size* and *Spacing* are non-zero the indifference point will not be at a 1:1 numerical ratio. In other words the participant can be biased, choosing one of two numerically equal arrays more than 50% of the time if that array has, for example, more spaced out dots.

The logarithmic model of choices in equation **(4)** is based on a particular hypothesis regarding the underlying internal representation of numerosity (Piazza et al. 2004). According to this hypothesis an approximate numerosity is represented as a normally distributed random variable. The distribution is centered on the actual value it is representing, but it is imprecise and probabilistic. The standard deviation of the numerosity random variable is the term *w*. When two numerosities are compared, as in a task like the one used here, the overlap in the two distributions causes confusability. Thus, the distance between the two numbers on the logarithmic mental number line (equivalent to

the ratio), and the *w* term determine the confusability of two numerosities and therefore the error rate.

The changes proposed to get from equation **(4)** to equation **(7)** correspond to an equivalent changes in the hypothetical underlying mental representation of numerosity. Instead of the numerosity normal random variable being centered on the actual number being represented, we propose that, in people who are biased by non-numerical features, the mean can vary depending on the *Size* and *Spacing* of the stimulus. Thus, the size and spacing of the items in a stimulus array can be thought of as increasing or decreasing the perceived numerosity depending on whether a particular participant has a positive or negative β_{Size} and $\beta_{Spacing}$.

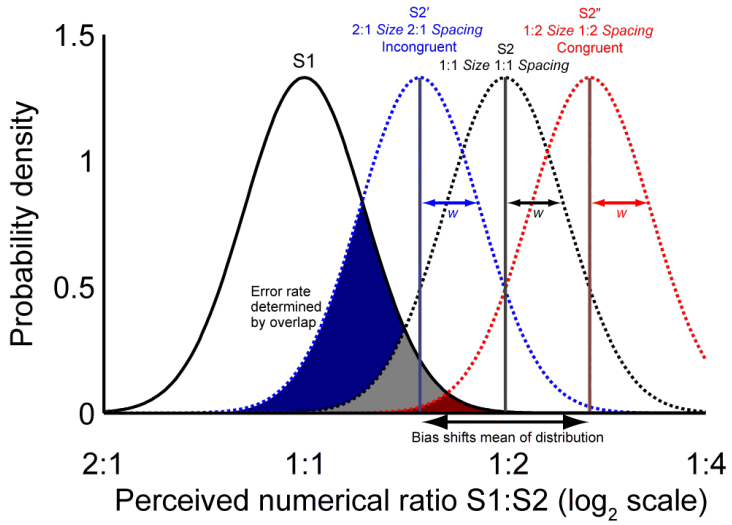

Figure 3-2. The internal representation of numerosity in a hypothetical participant according to the new model.
This hypothetical participant has the following coefficients: w = 0.3 ($\beta_{num} \approx 2.36$), β_{Size} = 0.5, $\beta_{Spacing}$ = 0.5, β_{side} = 0. Stimulus S1 and S2 are represented internally as normally distributed random variables on a logarithmically compressed mental number line. The standard deviation of these representations is fixed and equal to *w*. When *Size* and *Spacing* are equal in S1 and S2, the model is equivalent to the logarithmic model (equation 3-4; Piazza et al., 2010). However, when *Size* and *Spacing* are incongruent to number as in S2′ the perceived numerical ratio

decreases and confusability increases (as represented by the overlap of the distributions). When *Size* and *Spacing* are congruent (S2″) perceived numerical ratio increases and confusability decreases. The shift of the mean of the S2 distribution is given by equation **(10)**.

Figure 3-2 helps to elucidate the effects of *Size* and *Spacing* on the overlap of the internal representation of two numerosities, S1 and S2, in a hypothetical participant. If S1 and S2 differ by a numerical ratio of 1:2, but do not differ in *Size* and *Spacing*, then the model reduces to the logarithmic model (Piazza et al. 2010; Piazza et al. 2004) expressed in equation **(4)**. The overlap will depend only on *w* as illustrated by the S2 distribution in black in Figure 3-2. If the *Size* and *Spacing* ratios of S1 to S2 are both 2:1, however, the participant's bias causes the mean of the numerosity representation of S2′ to shift to the left as given by equation **(10)** (blue distribution). As a result, the overlap between S1 and S2′ increases and accuracy decreases, just as is actually observed in experiments in which non-numerical features are incongruent with numerosity. Conversely if the *Size* and *Spacing* ratios are 1:2, as in congruent trials, the distributions grow farther apart and accuracy improves (distribution S2″ in red). These changes in accuracy occur despite no change in numerical ratio and no change in *w*, and therefore cannot be modeled using previous approaches. Only a framework that takes non-numerical stimulus features into account can model these effects on error rate.

We have provided code in the supplementary materials that computes the *Size* and *Spacing* parameters and will fit the model in equation **(7)** to behavioral data sets.

3.3.7 The discrimination vector, discrimination dimension, and testing for non-numerical alternative strategies

The three value vector defined by β_{num}, β_{Size}, and $\beta_{Spacing}$ reflects the degree to which the distance between two stimuli along the three cardinal dimensions in Figure 3-1E affect the probability of choosing a particular stimulus as the more numerous one. We will refer to this vector as the participant's discrimination vector and the dimension it defines in Figure 3-1E stimulus space as the discrimination dimension. Pairs of stimuli that differ along the discrimination dimension are most easily discriminated, and participants are indifferent between pairs of stimuli that differ along the dimensions orthogonal to the discrimination dimension. If a participant has no bias, then her discrimination vector will be identical to the numerosity dimension, and the magnitude of the choice vector will be identical to β_{num}. However, if β_{Size} or $\beta_{Spacing}$ is not zero then the discrimination vector will differ from the numerosity dimension. A participant who has a significant β_{num} ($p < 0.01$) and no significant effect of β_{Size} or $\beta_{Spacing}$ ($p > 0.1$) can be considered to be making unbiased numerosity judgments.

Participants who fail to meet the criteria for unbiased numerical discrimination, may be primarily relying on numerosity but have a non-numerical bias or alternatively they may be responding primarily on the basis of one of many possible non-numerical stimulus features. Geometrically, this is equivalent to asking which of the named dimensions in Figure 3-1E is closest to the discrimination vector. To test this statistically we projected the discrimination vector onto the numerosity dimension (equal to β_{num}) and onto the each of the

other dimensions. Participants whose numerosity vector projection was significantly greater than all other vector projections ($p < 0.05$) were considered to be primarily relying on numerosity but biased by a non-numerical feature. Those whose numerosity vector projection was significantly smaller than another vector projection ($p < 0.05$) were considered to be primarily relying on a non-numerical strategy. Any participant whose numerosity vector projection was not significantly different from another vector projection were categorized as having an indeterminate response strategy.

3.4. Results

We fit the accuracy data of individual participants performing an ordinal approximate number discrimination task with choice curves with terms for side (left or right), guessing rate, numerical ratio, *Size* ratio and *Spacing* ratio. The *Size* and *Spacing* variables are defined mathematically in section 3.2.2. Intuitively, *Size* can be thought of as the aspect of the stimulus that changes with the size of a fixed number of items at fixed locations, and *Spacing* can be thought of as the parameter that changes when a fixed number of items of fixed size are spread out over a greater or lesser area of space.

3.4.1 Model fits account for performance variations due to non-numerical stimulus features

Figure 3-3A shows the model fit for the five participants who were tested with 7,500 trials. As the numerical ratio of items in the right array to items in the left array increased, participants became more likely to choose the right stimulus,

as instructed. The effect of numerical ratio on the probability of choosing “right” was well fit by the model across trials (black data points and fit lines in Figure 3-3A). In order to examine the effects of *Size* and *Spacing* and to evaluate how well the model accounted for these effects, we examined the subset of trials in which the non-numerical features differed dramatically. The red markers and green markers in Figure 3-3A reflect trials with large *Size* and *Spacing* ratios respectively (greater than a 8:3 or less than a 3:8 ratio). Critically, the model was only fit once for each participant to his or her full dataset; the red and green lines represent the predictions of the model for these subsets of trials. As can be visualized in the offset of the red and green lines from the black lines, all of these participants were influenced by *Size*, *Spacing*, or both. These red and green lines represent an explanation of variance in numerosity judgments that cannot be accounted for with either the logarithmic or linear models of the ANS used in previous studies.

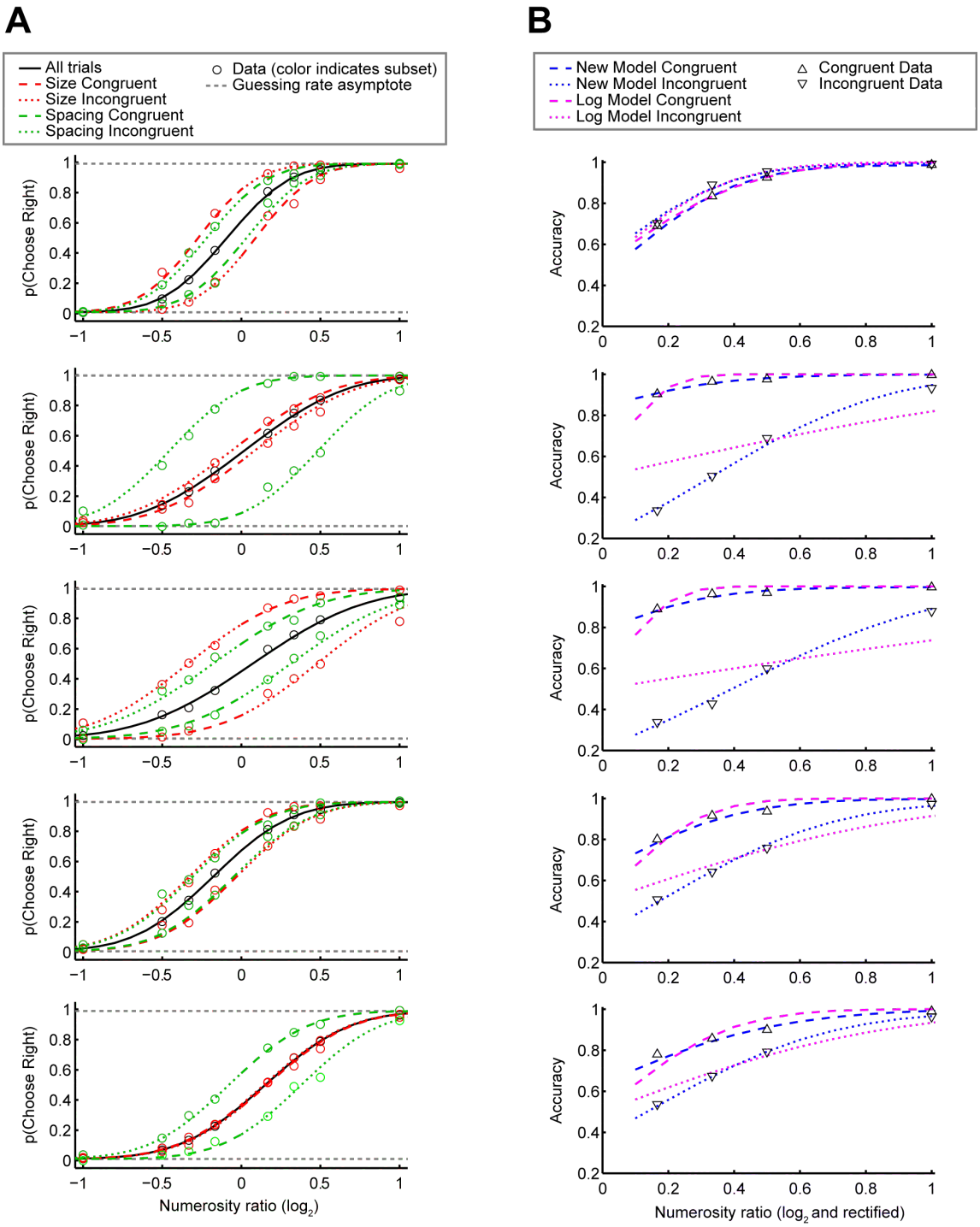

Figure 3-3. Modeling *Size* and *Spacing* explains congruity effects.
Each row of plots is a single participant's data. (A) Data (open circles) and model fit (black, red, and green lines). Dashed gray lines indicate model asymptote due to guessing rate (γ). The probability of choosing the stimulus array presented on the right is modeled as a function of the log of the left array to right array ratio of numerosity, *Size*, and *Spacing*. Black indicates the average of all data and the corresponding model fit. Red shows data and model fit for third of

trials with the greatest absolute ratio in *Size* and green shows the same for *Spacing*. Dashed lines indicate *Size* or *Spacing* was congruent with number and dotted lines indicate incongruent. All model predictions (lines) are derived from the parameters fit once to the entire dataset. (B) Data from the same participants plotted as accuracy. Upward pointing triangles indicate average data from all trials on which both *Size* and *Spacing* were congruent with numerosity, and downward pointing triangles indicate data from incongruent trials. Dashed lines are models fit only to congruent data points and dotted lines are fit only to incongruent data points. Blue lines were fit using the new model presented in this paper, and magenta lines using the standard logarithmic model based on Piazza et al. (2010; 2004).

On any given trial, the influence of *Size* or *Spacing* may help or hinder performance. If a participant has a significant positive effect of *Size* ratio or *Spacing* ratio, as do most of the participants in our sample, then larger and more spaced-out dots are perceived as more numerous (a notable exception is participant 1 who has a negative *Size* coefficient). As a result, when numerosity is congruent with *Size* or *Spacing*, a participant will be more likely to correctly identify the stimulus with the larger numerosity, as illustrated by the dashed red and green model fit lines in Figure 3-3A. In contrast, when *Size* or *Spacing* is incongruent, performance decreases as shown by the dotted red and green lines. In these trials, the larger and more spaced out dots make the less numerous stimulus appear more numerous and thus reduce accuracy. When the numerical ratio is difficult and the changes in *Size* and *Spacing* are large and incongruent to numerosity, participants can be induced to consistently incorrectly choose the less numerous stimulus at a rate greater than chance.

3.4.2 Modeling the effect of Size and Spacing improves w consistency across stimulus sets

One clear inadequacy of the current models of numerical discrimination is that they are incapable of modeling below chance performance. In particular,

when stimulus features are very incongruent to numerosity and the numerical ratios are difficult, the current models of numerosity discrimination will sometimes fail to converge, essentially estimating an absurdly large or infinite *w* (Szucs, Nobes, et al. 2013). We fit our model as well as the standard logarithmic (Piazza et al. 2010) and linear models (Pica et al. 2004) separately to just the congruent trials and just the incongruent trials to see if accounting for non-numerical feature bias helped reduce the variability in *w* estimates obtained from different stimulus sets. We considered a congruent trial as one in which the array containing more dots also had a larger *Size* and larger *Spacing* and incongruent as one in which the array containing more dots had the smaller *Size* and *Spacing*. As shown in Figure 3-3B, our model provides much better fits to the data than the standard logarithmic model (the fit of the linear model was not plotted because it overlapped so closely with the log model that it was difficult to distinguish). The fit is especially better on the difficult incongruent trials, on which some participants performed consistently below chance on the difficult numerical ratios.

Furthermore, the inter-condition reliability of *w* was higher and *w* more similar to the *w* obtained from fitting the full stimulus set for our model compared to the two other models. Figure 3-4 summarizes these results. When using the logarithmic and linear models, *w* obtained from incongruent trials tended to be much larger than for congruent trials, however *w* obtained from our model was quite similar across stimulus sets, with no discernable increasing or decreasing trend across the five participants. Thus, our model is capable of explaining large

differences in performance originating from non-numerical features. Furthermore, it properly attributes these performance effects to non-numerical model parameters with no systematic impact on the estimation of numerical acuity.

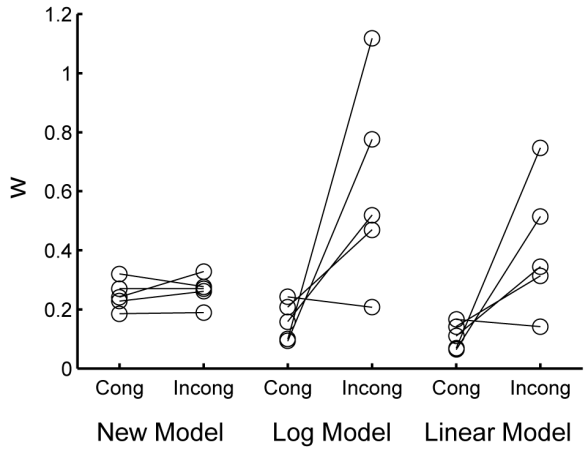

Figure 3-4. Modeling the effects of non-numerical features increases *w* reliability over changes in stimulus set.
The *w* coefficients calculated for the five participants who completed 7,500 trials by fitting the model presented in this paper, the logarithmic model (Piazza et al., 2004; 2010), and the linear model (Pica et al., 2004). Models were fit separately to just the data from congruent trials and incongruent trials.

3.4.3 Numerosity is the best explanation of performance but bias is universal

Figure 3-5 summarizes the effect of numerical ratio, *Size* ratio, and *Spacing* ratio on choice behavior for all of our 20 participants. Beta estimates are plotted in pairs as $\beta_{num} \times \beta_{Size}$, $\beta_{num} \times \beta_{Spacing}$, and $\beta_{Size} \times \beta_{Spacing}$ in Figure 3-5A – 3-5C respectively. The estimates are plotted as standard error ellipses to denote the confidence of the estimate. The small ellipses represent more precise beta estimates derived from the five participants who performed 7,500 trials (some errors are so small they may appear as points), and larger ellipses reflect the 15 participants who performed 750 trials each. These three beta estimates comprise the discrimination vector of each participant. The direction of the

discrimination vector (the discrimination dimension) represents what stimulus features a participant is utilizing to make her choices, and the magnitude of the discrimination vector represents the participants' acuity in discriminating that feature (see section 3.3.7 for further explanation).

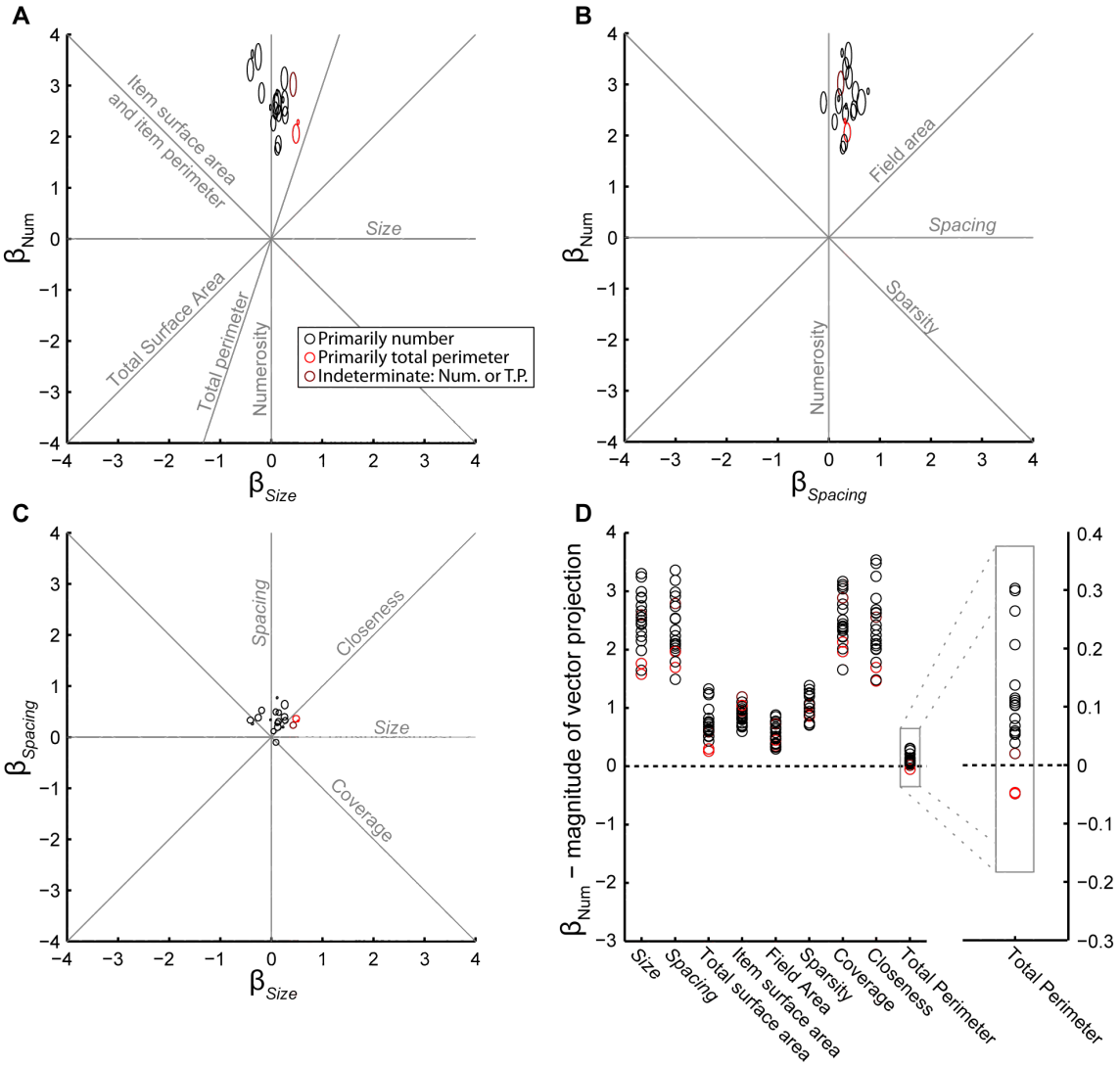

Figure 3-5. No single feature fully explains choice behavior in any participant, but numerosity is the closest approximation for 18 out of 20 participants.
(A – C) Estimates of effect sizes of the log of the ratio of numerosity, *Size*, and *Spacing* plotted in pairs for all 20 participants represented as standard error ellipses. Black ellipses indicate the 17 participants whose choices were better explained by numerical ratio than by any other stimulus feature, red indicates the two better explained by total perimeter, and dark red the one participant who may have been either choosing on the basis of number or total perimeter. Stimulus feature

dimension lines are shown as gray lines. The alignment of β estimates (discrimination vector) with a feature dimension is indicative that that feature is driving choice behavior. (D) The magnitude of the projection of the discrimination vector onto each non-numerical dimension subtracted from the magnitude of the projection onto the numerosity dimension. Positive values indicate that number was a better explanation of choices than the feature it is being contrasted against. Negative values indicate that the non-numerical feature was a better explanation of choices. Colors indicate the same participant groups as above. Axis on the right side is the same data from the contrast between number and total perimeter expanded for easier inspection.

The hypothesis that numerosity is the sole factor driving behavior is equivalent to the hypothesis that the beta for numerosity is significantly different from zero, and *Size* and *Spacing* betas are not different from zero. Although the choices of all of the participants in our sample were significantly influenced by numerosity ($p << 0.001$), none of them met our criterion for "pure" numerosity discriminators. In other words, we could not rule out the possibility that *Size* and *Spacing* might also be influencing behavior ($p < 0.1$ for all participants), and so we categorized them as "biased". Thus, we could rule out what we consider to be the implicit hypothesis of the two dominant models of numerosity discrimination: that the ratio of the numerosities and ANS acuity (as estimated by *w*) are the only factors affecting numerosity discrimination performance.

The beta space represented in Figure 3-5 is analogous to the stimulus space in Figure 3-1E. In particular, the log ratios of the non-numerical stimulus features can be expressed as linear combinations of the log ratios of numerosity, *Size*, and *Spacing*. Thus, the hypothesis that a particular stimulus feature contributes to choice behavior is equivalent to the hypothesis that particular linear combinations of log numerosity, *Size*, and *Spacing* shape choice behavior.

Those linear combinations are represented as “feature” dimensions in Figure 3-5A – 3-5C and labeled with the stimulus feature to which they correspond.

This produces a simple graphical representation of the features driving individual participants’ choice behavior. For example, if we had found a participant who made choices based only on the number of dots while ignoring the dots’ size and spacing, the discrimination vector error ellipse for that participant would lie on the numerosity feature dimension in both Figure 3-5A and 3-5B and at the origin in Figure 3-5C. Alternatively, if a participant always and only relied on total surface area to discriminate stimuli, her beta parameters for numerosity and *Size* would be significantly positive and equal, but the parameter for *Spacing* would be near zero. The exact numerical values of the numerosity and *Size* beta parameters would indicate the participant’s acuity in discriminating total surface area. As a result that participant’s discrimination vector would fall along the “total surface area” dimension in Figure 3-5A. The slope of each of these feature dimension lines in Figure 3-5 is determined by the linear equations in Appendix A. Geometrically, asking which stimulus feature is determining behavior is equivalent to asking which feature dimension a participant’s discrimination vector lies on. If the discrimination vector is significantly offset from all features lines, we may ask which feature *best* explains behavior. This is the equivalent of asking to which feature dimension the discrimination vector is closest.

To determine if a participant's discrimination vector was significantly closer to a non-numerical feature dimension or to the numerosity dimension, we projected the discrimination vectors onto each feature dimension. We then tested whether β_{num} was significantly larger than all the other vector projections using linear contrast hypothesis testing. If a particular vector projection was significantly larger than β_{num} we could conclude that the participant was more influenced by that parameter than by numerosity. Figure 3-5D shows the difference between β_{num} and the magnitude of the discrimination vector projected onto each of the other stimulus features. Two out of twenty participants were significantly better described as basing choices on total perimeter rather than numerosity ($p = 0.009$ and $p << 0.001$ for linear contrast). One additional participant could not be categorized and was either a numerosity or a total perimeter discriminator ($p = 0.31$). All 17 other participants were significantly better described as discriminating numerosity than as discriminating any other stimulus feature ($p < 0.05$).

3.4.4 Advantages of new model hold with fewer trials

We ran our participants on many trials to ensure that we were able to precisely quantify their response strategies and bias terms. However, most studies of numerical cognition run fewer trials. If we hope for broader adoption of our modeling approach then it would be useful to know if the main advantages outlined here apply when fewer trials are used. We reran our analyses on reduced numbers of trials for all participants. Figure 3-6A shows the proportion of

the 20 participants that met the criteria outlined above for the effect of non-numerical features. We tested for bias: that some factor besides numerosity affects choices, and for strategy: that numerosity or some other factor is the primary determinate of choices. Although the ability to detect bias and the primary determinate feature decreases with fewer trials, both measures were effective to as few as 250 trials. It is worth noting that a study designed with larger *Size* and *Spacing* ratios would be more sensitive to bias using even fewer trials.

Figure 3-6B recapitulates the finding in Figure 3-4 for all participants and using fewer trials. Inter-method reliability of *w* is higher using our new model than for either the logarithmic or linear models both of which over estimate *w* when non-numerical stimulus features are congruent with number. On average this is true regardless of the number of trials, but the variance in the ratio increases with fewer trials.

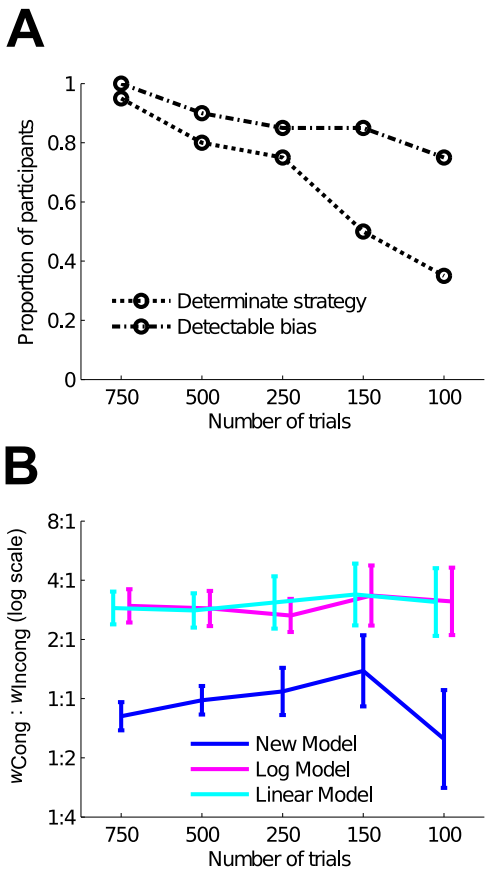

Figure 3-6. The advantages of the new model are evident even with fewer trials. (A) The proportion of participants for whom one of the ten features explains the significantly greatest portion of the variance (determinate strategy) is plotted against the number of trials used in the analysis. The proportion of participants in whom we could detect bias is also plotted against number of trials. (B) Extending the analysis in Figure 3-4 to all participants, *w* was calculated separately for congruent and incongruent trials (w_{Cong} and w_{Incong} respectively). The mean of the ratio of w_{Cong} and w_{Incong} is plotted on a log scale for the three models. Error bars denote standard errors of the means (n = 20 participants).

3.5 Discussion

The model for estimating *w* presented here and the stimulus space on which it is based represents an advance over previous approaches in four important ways. First, the stimulus space itself identifies, for the first time, the three degrees of freedom available to ANS researchers in designing stimuli, and elucidates the tradeoffs and partial collinearities inherent to arrays of dots.

Second, the model reconciles the concept of numerosity as an internal random variable on a log scale mental number line with the fact that non-numerical features also affect discrimination performance by allowing the mean of the random variable to shift with changes in the size and spacing of the dots within the stimulus arrays. Third, by correctly attributing correct and incorrect responses caused by congruence or incongruence of non-numerical features to the *Size* and *Spacing* parameters, the model yields a *w* that is a more valid estimate of numerical acuity and is more reliable over different stimulus sets. Fourth, our model provides an alternative approach for assessing the role of non-numerical stimulus features on numerosity judgments. Rather than attempting to control for non-numerical features by equating different dimensions in different trial subsets, our approach is to intentionally vary non-numerical features and to model their effects on performance.

Our model applied to the data set presented here demonstrates that number is all participants were influenced, to some extent, by non-numerical features while attempting to perform a numerical discrimination. Although the effect of non-numerical features on numerical estimation and comparison has been well documented in the literature, our data show that these effects are nearly universal even among educated adults. Our data further shows that for at least 17 out of 20 subjects, out of the comprehensive list of ten stimulus features tested, number best explained behavior. We consider this strong evidence that

numerosity does exist as an internal magnitude, and that it cannot be explained away as “merely” the derived effect of other features.

3.5.1 Stimulus space and modeling

The two dominant models of numerosity discrimination do not adequately account for the effects of non-numerical stimulus dimensions on accuracy. Both the logarithmic (Piazza et al. 2010; Piazza et al. 2004) and linear (Halberda, Mazzocco, and Feigenson 2008; Pica et al. 2004) models posit that numerosity is internally represented as a distribution or random variable along a mental number line, with a mean equal to the number represented. The width of the distribution may be either fixed (log model) or vary with the magnitude being represented (linear model). In either case *w* is proposed to be a measure of the fuzziness of the internal representation of number intrinsic to the individual. Critically, both models posit that performance in a numerical ordering task is determined only by *w* and the numerical ratio. Empirically, however, many groups have demonstrated that non-numerical stimulus features do indeed affect performance in numerical ordering tasks (e.g. DeWind and Brannon 2012; Frith and Frith 1972; Gebuis and Gevers 2011; Ginsburg 1976; Sophian 2007; Tokita and Ishiguchi 2010).

Here we extend the standard logarithmic model of numerosity perception and discrimination to include terms that capture the effects of the size and spacing of the dots in the stimulus arrays. In our new revised logarithmic model, numerosity is represented as a random variable on a log compressed mental

number line. However, the size and spacing of the dots in the stimulus array can cause the mean of this distribution to be shifted to a position greater or less than the actual number of items in the stimulus. As a result, the overlap of two numerosity distributions, and therefore the predicted error rate, may be larger or smaller depending not only on numerical ratio but also on whether non-numerical features are congruent or incongruent with numerosity (see Figure 3-2 for a hypothetical example). By accommodating the effects of non-numerical features, our model is able to capture variance in numerical discrimination behavior that went unaccounted for in previous models.

Extending the logarithmic model of numerosity, however, is not as straightforward as simply adding regressors for each non-numerical feature that might influence numerical perception. Such a model would be overdetermined due to the partial collinearity of these features. It was therefore essential to identify the mutually independent regressors that fully describe the stimulus features that could affect performance on a numerosity discrimination task.

Relying heavily on the framework of intrinsic and extrinsic features pioneered by Dehaene, Izard, and Piazza (2005), we developed a novel stimulus space. For the first time we provide a comprehensive description of dot array stimuli that encompasses the critical features affecting numerical discrimination. This space has three dimensions that describe the number, size, and spacing of the dots in an array, a formulation that is complete but not redundant. This space provides ANS researchers with a powerful new tool for understanding the

tradeoffs and collinearities inherent to dot array stimuli, as well as providing the basis for quantifying the effects of ten different stimulus features on numerical discrimination.

3.5.2 A more valid and reliable w

Is Weber fraction (*w)* a valid measure of nonverbal numerical acuity? In previous studies, estimates of *w* have failed to account for the effect of non-numerical features in a systematic and quantifiable way. As a result, they implicitly assume that all errors result from imprecision of the internal numerosity representation. However, we found that differences in non-numerical features affected perceived numerosity in all twenty of our participants. In contrast to the logarithmic and linear models used in previous studies which yield *w* estimates that are an amalgamation of the effects of number and non-numerical features, our model allows an estimate of *w* that is independent of the effects of Size and Spacing. In this sense, *w* derived from our model is a more valid measure of numerical acuity.

The practical corollary of a more valid measure of numerical acuity is an increase in “inter-method” or “alternate-form” reliability. This type of reliability refers to the tendency of different tests to generate the same result. In this case the different tests of numerical acuity are different stimulus sets that vary non-numerical features in different ways. To assess this type of inter-method reliability we compared the w estimates for the new revised logarithmic model to the two standard models for congruent and incongruent trials separately. For

both standard models, *w* estimates were much higher in the incongruent compare to the congruent condition. In contrast, our new revised logarithmic model returned similar estimates for the two trial types and therefore showed more inter-method reliability. The discrepancy in w estimates for incongruent and congruent trials observed under the standard models has been observed previously, and has been interpreted to mean there is no stable internal representation of numerical magnitude, and has even been offered as evidence against the existence of the ANS (Szucs, Nobes, et al. 2013). Our model demonstrates that *w* is in fact stable over these stimulus conditions and that the instability observed in previous studies was due to the fact the estimates of *w* were not independent Size and Spacing. Future work should compare these models across laboratories, stimulus sets, and task designs to more fully assess reliability.

There are several potential benefits of a more reliable, valid, and cross-paradigm comparable measure of numerical acuity. Recently, there has been interest in the predictive power of numerical acuity on mathematical achievement (DeWind and Brannon 2012; Gilmore et al. 2013; Gilmore, McCarthy, and Spelke 2010; Halberda et al. 2012; Halberda, Mazzocco, and Feigenson 2008; Lyons and Beilock 2011; Park and Brannon 2013; Piazza et al. 2010; Starr, Libertus, and Brannon 2013; Mazzocco, Feigenson, and Halberda 2011a). These correlations, however, are relatively weak and only predict a small amount of variance in mathematical performance. Some researchers have also suggested

that non-symbolic numerical abilities are part of a larger suite of visual-perceptual abilities that predict mathematics performance (Tibber et al. 2013). Others have argued that ANS acuity provides unique variance to predicting mathematical performance and that other similar perceptual tasks do not (Agrillo, Piffer, and Adriano 2013), or that both ANS acuity and other perceptual tasks provide unique variance (Lourenco et al. 2012). Parsing out non-numerical bias from numerical acuity may improve these correlations by reducing the effect of bias on *w* estimates. Alternatively, bias itself might be a mediating factor. Participants who cannot clearly differentiate numerosity from other magnitudes may have impaired performance on other perceptual tasks or with symbolic mathematics itself. For example, some have suggested that the "stroop like" aspect of numerosity discriminations with strong non-numerical feature incongruity reveals difficulties inhibiting prepotent responses (Fuhs and McNeil 2013; Szucs, Devine, et al. 2013). They suggest that previous correlations between *w* and math achievement may be mediated by failure to inhibit responses to other stimulus features. Isolating bias from numerical acuity will allow these hypotheses to be tested more directly.

3.5.3 Stimulus Control

The goal of stimulus control in numerosity experiments has been to ensure that numerosity is driving choice behavior. In the literature there are two standard ways of accomplishing stimulus control in dot array comparison tasks. Both of these methods have drawbacks, and neither gives an objective measure

of non-numerical feature bias. Our modeling approach, in contrast, provides a clear quantitative measure of both numerical acuity and bias and can detect alternative response strategies that are based primarily on non-numerical features of the stimulus.

The most common method for controlling non-numerical features is to divide trials into sets that each control for a different non-numerical stimulus feature, an approach adopted by many research groups (Ansari and Dhital 2006; Halberda, Mazzocco, and Feigenson 2008; Izard et al. 2009; Libertus, Woldorff, and Brannon 2007; Piazza et al. 2010; Santens et al. 2010). For example, if total surface area were fixed in one set of trials, numerosity and item surface area would vary together. In another set of trials, item surface area would be fixed and total surface area would vary with numerosity. If a participant were relying on one of these features as a proxy for numerosity then choice behavior would be at chance on the subset of trials on which that feature is fixed. This basic logic certainly works for ruling out total reliance on a particular feature; however, the analysis is underpowered since it relies on a subset of trials. Furthermore, for practical reasons most studies do not control all possible parameters in different trial subsets; a problem which is particularly salient given our finding that total perimeter, a rarely controlled parameter, is subserving a non-numerical strategy in some people.

Another common approach to stimulus control is to have subsets of trials in which a particular non-numerical stimulus feature is varied in a manner either

congruent with or incongruent with numerosity (e.g. Cantlon and Brannon 2005; DeWind and Brannon 2012; Hurewitz, Gelman, and Schnitzer 2006; Rousselle and Noël 2008; Szucs, Nobes, et al. 2013). Ruling out a non-numerical strategy is particularly problematic using this paradigm. It depends on observing above chance performance in incongruent trials, but as can be seen in Figure 3-3B, participants with any bias at all can be induced to consistently choose the incorrect stimulus when numerical ratios are very difficult and the ratio of incongruent non-numerical features is very large. Thus, the test for non-numerical strategies is too sensitive and can interpret small effects of non-numerical features as total reliance on them (Szucs, Nobes, et al. 2013). The sensitivity of the test depends on paradigm idiosyncrasies such as the difficulty of the numerical ratios and the degree of variation in non-numerical features. Furthermore, like the first method mentioned above, a design that attempts to control for all non-numerical stimulus features using a congruent-incongruent paradigm would require a multitude of conditions (e.g., perimeter congruent, surface area incongruent, etc.). A recent paper claimed to have resolved the problem of stimulus control for dot arrays using this congruent-incongruent approach (Gebuis and Reynvoet 2011). Unfortunately, their approach suffers from the same intrinsic problem outlined above.

A critical insight derived from our stimulus space and modeling approach is that non-numerical bias, which we define as the marginal effect of a non-numerical feature on choices, and a non-numerical strategy, which we define as

the primary reliance on non-numerical features, exist on a continuum. By varying numerical and non-numerical stimulus features and modeling their effect on choices, our paradigm provides a quantitative measure of non-numerical feature bias. If these terms are sufficiently large, then choices will be better described by the ratio of a non-numerical feature than by the ratio of numerosity itself, and we consider such a participant to be utilizing a non-numerical strategy. Furthermore, these analyses are made based on the entire dataset, not subsets of trials, and therefore have more statistical power.

This study represents the most comprehensive effort of which we are aware to simultaneously quantify the effect of as many non-numerical features as possible on the internal representation of number. Thus, although bias was universal among our participants, it is worth noting that 17 out of 20 participants used numerosity more than any other feature to make their discriminations. We take this as evidence that numerosity is not reducible to “merely” the effects of other stimulus features as suggested by some (Gebuis and Reynvoet 2012a; Gebuis and Reynvoet 2012c; Szucs, Nobes, et al. 2013), but is itself an important determinant of behavior. Two of the twenty participants, however, had such large *Size* bias that it was more parsimonious to describe them as discriminating total perimeter than as discriminating numerosity. This finding makes the importance of controlling total perimeter in ANS studies apparent.

We also found a large and relatively consistent effect of *Spacing* on numerosity judgments. Nineteen out of twenty participants viewed arrays with

more spaced out dots as more numerous. This effect has been noted before (Allïk and Tuulmets 1991; Dakin et al. 2011; Kramer, Di Bono, and Zorzi 2011), and may provide some insight into the processes by which numerosity is extracted from the visual scene.

3.5.4 Approximate Number System and/or an Approximate Magnitude System?

We use the term ANS throughout this paper because we believe that it is a useful construct however it is important to emphasize that our model and data set are not designed to test for the existence or lack thereof of an ANS. A recent study purported to provide evidence against the existence of an ANS based on low within subject reliability for stimulus sets for which non-numerical variables were congruent or incongruent with number (Szucs, Nobes, et al. 2013). However, as we explained in sections 3.5.2 and 3.5.3 the low reliability they obtained can be attributed to failing to model non-numerical features. Thus their data do not address the existence or lack thereof of a dedicated system for representing number approximately.

One of the advantages of our stimulus space and model is that it illustrates the close relationship between number and other features of the stimulus. In Figure 3-5A – 3-5C especially, it is clear that a small effect of *Size* or *Spacing* can be considered a marginal biasing effect on numerical discrimination, but a sufficiently large effect is better described as an alternative response strategy (albeit likely an unconscious one). This continuum of effects may lend support to the idea that rather than an ANS there is a more general approximate

magnitude system that allows approximate enumeration, but also subserves approximation of other continuous properties of a stimulus. From this perspective, our findings can be seen as supporting Walsh's theory of magnitude (ATOM), which suggests that all magnitudes share a common currency, or at least overlapping representation in the brain (Bueti and Walsh 2009; Cantlon, Platt, and Brannon 2009; Walsh 2003).

A related question is how the representation of number and continuous variables emerges over human development. One possibility is that numerosity is conflated with other magnitudes early in development, but that over development numerosity becomes more differentiated (Lourenco and Longo 2010; Walsh 2003). Within the context of our model, confusion of different stimulus dimensions would manifest itself in the magnitude of the*Size* and *Spacing* coefficients. The classic Piagetian view is that early in development children attend to size and volume and only later come to appreciate number as an abstract variable. Seemingly consistent with this view a handful of studies found that perimeter or area are more readily encoded by infants than number (Clearfield and Mix 1999; Clearfield and Mix 2001; Feigenson, Carey, and Spelke 2002) However, other data is inconsistent with this view and suggests that infants spontaneously encode both kinds of information. For example, Libertus, Starr and Brannon (2014) used a visual change detection paradigm and found that when infants were shown two streams of visual images where one stream alternated numerically and the other alternated in total surface area infants

preferred to look at the numerically changing stream (Cordes and Brannon 2009). We hope that our model can be used to assess the relative strength of the number vector vs. vectors for continuous variables in young children's decision-making and to track changes in numerical sensitivity and bias over development.

Relatedly, comparative studies of other species have examined relative use of number and other features. Monkeys and many other animal species can be trained to attend to number and largely ignore other visual features (Brannon and Terrace 1998; Cantlon and Brannon 2006). It has been suggested, however, that this ability is not part of animals natural behavioral repertoire and only results from extensive training (Seron and Pesenti 2001). Cantlon and Brannon (2007) offered evidence against this view. They trained rhesus monkeys to match stimuli based on numerosity and a redundant non-numerical variable such as color, shape, or surface area. Once monkeys reliably matched these redundant cue stimuli, they were given a choice between one stimulus that matched the sample numerically and another stimulus that matched based on the previously redundant variable (e.g., color, shape, or surface area). The monkeys decisions were strongly influenced by the numerical distance between the sample and incorrect numerical match (Cantlon and Brannon 2007). Furthermore this was true even for one monkey who had no prior numerical training. In contrast research with some other species such as mosquito fish suggest that their quantitative judgments may be more influenced by continuous variables (Agrillo, Piffer, and Bisazza 2011).

Thus, the preponderance of recent developmental and comparative evidence suggests that number is more than a “last resort” strategy for disambiguating stimuli. We hope that our model can be used to quantify the role of number vs. non-numerical variables on behavioral decisions and be use to study species differences as well as changes as a function of training and experience.

3.5.5 Future Directions

An important future direction is to see how well bias can be estimated in previously collected published datasets and to see how these bias estimates, as well as the new estimates of numerical acuity that account for bias related errors, change or clarify previous hypotheses. Our modeling approach does not depend on a particular esoteric arrangement of stimulus parameters. We orthogonalized number, *Size*, and *Spacing* ratios to increase power, however as long as these features are not perfectly collinear our modeling approach can be applied to data sets acquired using diverse stimulus control paradigms. To facilitate the adoption of our model we have included computer code in a supplement to this research article.

An important advantage inherent to our model of choice behavior is that it easily accommodates more regressors to model other important aspects of choice behavior. We argue that the advance made in this paper is the observation that the continuous parameters thought to affect numerical discriminations can be reduced to three regressors that can be varied

independently. A fourth regressor was added to account for side bias. More regressors could be added for other variables that can be varied independently from numerosity, *Size*, and *Spacing*, for example, brightness, contrast ratio, or item shape.

The standard way of measuring *w* is to present pairs of dot arrays and require participants to make an ordinal judgment. The number of parameters that can freely vary between research groups without being expressly modeled is shrinking. Here we modeled the effects of *Size* and *Spacing* for the first time and by extension all of the derived features in Appendix A. Our model, however, cannot explain some other features known to affect the perception of numerosity. First, our model does not account for the effect of stimulus exposure time. Inglis and Gilmore (2013) demonstrated that stimulus exposure time is a critical variable that must be accounted for when estimating *w* and they provide a model for doing so. Second, although our model contains a term for item spacing, it cannot account for the effect of items "clumping" within the array. The solitaire illusion (Frith and Frith 1972) and the regular-random illusion (Ginsburg 1976) demonstrate that clumping does affect numerosity estimates. The occupancy model (Allïk and Tuulmets 1991) provides a modeling framework for explaining these effects. Other effects that our model does not address are the effect of the absolute magnitude of the values being compared separate from ratio (Prather 2014),and hysteresis whereby the difficulty and perceptual qualities of the previous trial effects current discrimination (Cicchini, Anobile, and Burr 2014;

Odic, Hock, and Halberda 2014).Integrating the effects of these visual features into our model is beyond the scope of this paper. However, future work should explore the interactions between exposure time, clumping, absolute magnitude, hysteresis and numerical acuity and bias to further reconcile different paradigms, aid in comparisons across paradigms and research groups, and deepen our understanding of the mechanisms of approximate enumeration.

Although most investigations into the ANS use static arrays of dots, similar stimulus control problems exist for aural or visual numerical stimuli presented sequentially. The extrinsic variables analogous to total surface area and field area would be total event duration and total stimulus duration respectively. The intrinsic variables analogous to item surface area and sparsity would be individual event duration and mean event period (equivalently total stimulus duration per event or the reciprocal of frequency). Numerosity-independent variables analogous to *Size* and *Spacing* could be generated by the same equations and a regression model closely analogous to the one presented here could be adopted.

We focused on choice behavior in this study, however the stimulus space and model could be used on any dependent variable that might vary with *Size*, *Spacing* and numerosity. For example, various studies have looked at the effects of dot array numerosity on BOLD signal (Cantlon et al. 2006; Jacob and Nieder 2009; Piazza et al. 2004; Piazza et al. 2007), EEG (Gebuis and Reynvoet 2012b), and the firing rates of individual neurons (Nieder and Miller 2004a;

Roitman, Brannon, and Platt 2007). Currently, non-numerical features are treated as nuisance variables that must be controlled. Our approach of quantifying non-numerical features allows the stimulus space dimensions affecting neurological dependent variables to be teased apart. We hope that applying similar modeling approaches to the one used here will lead to a better understanding of how low level visual features processed early in the cortical visual stream are transformed into the numerosity signals seen in the intraparietal sulcus and prefrontal cortex.

Finally, we anticipate that this model will be useful for looking at changes in the salience of non-numerical features over development and individual differences in the influence of non-numerical variables on numerical discrimination at a given age.

3.5.6 Conclusions

We extended the logarithmic model of numerical acuity to dissociate the biasing effects of Size and Spacing from *w*. Instead of merely controlling for non-numerical stimulus features the model allows a quantification of the effect of non-numerical stimulus features on choices. The model applied to our data set demonstrates that non-numerical features widely affect numerical discriminations in adults but that for most individuals these effects are relatively small compared to the effect of number itself.

4. A population code for visual magnitude in macaque ventral intraparietal area

4.1 Introduction

The “number sense” describes our intuitive concept of numerical quantity, a feeling we have for the number of things we can see or hear that does not require counting (Dehaene 1997). Educated adult humans share this ability to enumerate approximately with infants (e.g. Xu and Spelke 2000) including neonates (Izard et al. 2009), non-human primates (e.g. Brannon and Terrace 1998; Tomonaga and Matsuzawa 2002), and other vertebrates (e.g. Honig and Stewart 1989; Meck and Church 1983). Furthermore, numerical intuitions exist without explicit training on numerical tasks in primates (Hauser et al. 2003; Lewis et al. 2005; Santos et al. 2005), and although some human languages lack words for individual numbers, individuals who have never been exposed to number words or numerals nevertheless can enumerate approximately (Pica et al. 2004). In light of these findings, the number sense has been described as a “core system of numerical representation”, because it is pre-linguistic and requires no cultural transmission or learning (Feigenson, Dehaene, and Spelke 2004).

Recently, researchers have made great strides in elucidating the neurobiological basis of the number sense. Functional imaging techniques have revealed that the horizontal segment of the intraparietal sulcus (hIPS) is activated

during approximate calculation (Dehaene et al. 1999), and that blood-oxygen-dependent signal recorded from hIPS is modulated by the number of items in a visual array even in the absence of any numerical task (Piazza et al. 2004; Jacob and Nieder 2009). The same areas also respond to number words and written numerals, indicating that human hIPS represents the abstract numerical magnitude underlying symbols (Piazza et al. 2007; Eger et al. 2003). Single neuron recordings in macaque intraparietal sulcus (IPS) and dorsolateral prefrontal cortex (DLPFC) have demonstrated that neurons modulate their firing rate in response to the number of items in a visual array (Nieder and Miller 2004a; Roitman, Brannon, and Platt 2007).

Single neuron recordings in monkeys and computational modeling of how number might be extracted from the visual scene have helped elucidate the spiking code for numerosity. Many theoretical and neural network models of visual and abstract numerosity exist (Dehaene and Changeux 1993; Verguts and Fias 2004; Verguts and Fias 2008; Whalen, Gallistel, and Gelman 1999; Zorzi and Butterworth 1999). All of these networks share a layer that can be understood as a “summation code” with units that increase their activity monotonically with numerosity. Roitman et al. (2007) recorded single neurons that both increased and decreased monotonically with numerosity in the lateral intraparietal area (LIP), providing neurobiological evidence for a summation layer. Some of these models (Dehaene and Changeux 1993; Verguts and Fias 2004; Verguts and Fias 2008) also posit a “labeled line code” network layer,

which is marked by units that are each tuned to a preferred numerosity that elicits maximum firing, while other numerosities elicit lower firing rates that attenuate as a function of numerical distance from the preferred numerosity. Single neurons following this pattern of activity have been found in DLPFC and the IPS with most such neurons concentrated in the fundus of the sulcus, the ventral intraparietal area (VIP; Nieder and Miller 2004a). It is hypothesized that these neurons are the neural instantiation of the labeled line layer in the neural networks, and that this "output layer" forms the neurobiological basis of mental representation of number upon which both the subjective sense of number and numerical estimations are founded (Nieder and Dehaene 2009).

Given the confluence of evidence demonstrating that the number sense is prelinguistic and independent of culture and that the IPS contains two types of representations thought to be fundamental to the extraction of numerosity from the visual field, it is natural to assume that the IPS is biologically specialized for numerical processing. This hypothesis is strongly supported by recent evidence of tuned numerosity neurons in the IPS of monkeys that have never been trained on a numerical discrimination task (Viswanathan and Nieder 2013).

Of course, evidence that the IPS contains numerosity signals does not necessitate that the IPS is uniquely specialized for number. There is a large body of evidence demonstrating that IPS represents many other visual stimulus features and has other non-visual response properties. Lateral intrapariatel area (LIP) is known to play a critical role in covert visual attention, saccade planning,

and saccade initiation (Bisley and Goldberg 2003; Colby and Goldberg 1999; Colby, Duhamel, and Goldberg 1996) and has more recently been implicated in abstract category representation (Freedman and Assad 2006). LIP also contains information about shape (Sereno and Maunsell 1998) and motion (Fanini and Assad 2008). A critical and unifying aspect of LIP is that it alters its response based on the behavioral relevance and expected value of a stimulus (Kiani and Shadlen 2009; Platt and Glimcher 1999; Shadlen and Newsome 2001).

VIP is an anatomically and functionally distinct region of the IPS (Colby, Duhamel, and Goldberg 1993) characterized by strong motion direction tuning including cells tuned to complex motion stimuli such as optic flow fields (Bremmer, Duhamel, et al. 2002). It is also a multimodal region with tactile, vestibular, auditory, and visual responses (Bremmer, Klam, et al. 2002; Duhamel, Colby, and Goldberg 1998; Schlack 2005). This pattern of response properties has been theorized to comprise a representation of self-motion during three-dimensional navigation (Bremmer, Klam, et al. 2002) or of approaching stimuli in peri-personal space (Graziano and Cooke 2006).

There is also evidence suggesting that parietal cortex and the IPS in particular is involved in the representation and comparison of other magnitudes such as time, space, and size (for reviews see Bueti and Walsh, 2009; Cantlon et al., 2009; Hubbard et al., 2005; Walsh, 2003). Similar psychophysical functions when comparing non-numerical magnitudes suggest common comparative processes, and interference has been observed between magnitude dimensions

when participants are asked to perform dual tasks (Dormal and Pesenti 2012). Functional imaging studies demonstrate that when humans compare magnitudes such as spatial location, luminance, and size, overlapping areas of the intraparietal sulcus are recruited (Cohen Kadosh and Henik 2006; Pinel et al. 2004; Zago et al. 2008). Single cell recordings in monkeys by Tudusciuc and Nieder (2007) have demonstrated that both numerosity and line length are encoded in IPS in partially overlapping populations. Thus, one possibility is that the intraparietal sulcus is specialized for the comparison of magnitudes, as opposed to the representation of numerosity.

In light of the evidence that the IPS encodes a more general representation of visual stimuli and the fact that IPS neurons are known to alter their tuning depending on task demands, we tested the hypothesis that neurons in the IPS might broadly represent many visual magnitudes before an animal is trained on an explicit numerical discrimination task. We used a stimulus set in which many visual features varied independently of numerosity and combined this with a novel mathematical description of dot array stimuli to simultaneously examine the role of number and ten non-numerical visual features of dot array stimuli in driving the firing rate of single neurons in VIP. When we examined which stimulus features drove changes in neural firing rate, we found that some neurons were indeed modulated by number and no other visual feature. However, numerosity selective neurons were few, and, like most cortical

representations, were noisy. A larger proportion of neurons were selective for non-numerical stimulus features.

We also used a simple classifier to try to estimate stimulus parameters from firing rate. We found the numerosity of the stimulus was better predicted from the entire population of feature-selective neurons than from the small group of neurons modulated by only numerosity. This finding suggests that the number sense, at least in untrained animals, may not rely on individual neurons tuned to specific numerosities. Instead, our modeling suggests that numerosity information may be extracted from the entire population of neurons by r

4.2 Results

We recorded 118 neurons in VIP from two monkeys. Monkeys fixated centrally while we presented arrays of 2, 4, 8, 16, or 32 dots within the predetermined visual receptive field of the neuron being recorded. Monkeys received a juice reward for fixating through the entire trial irrespective of the stimulus features.

Figure 3-1A shows a peri-stimulus time histogram of one neuron sensitive to changes in numerosity. The cell is typical in showing a phasic response to the onset of the stimulus and a more slowly emerging differentiation by stimulus condition. The inset shows the average response of the same neuron across numerosities in the stimulus epoch, showing that as numerosity increases, the firing rate decreases. Figure 3-1B shows the peri-stimulus time histogram of another neuron that also appears to be sensitive to numerosity. The stimulus

epoch average in the inset shows that the cell responds non-linearly to changes in number and responds maximally to 32 dots.

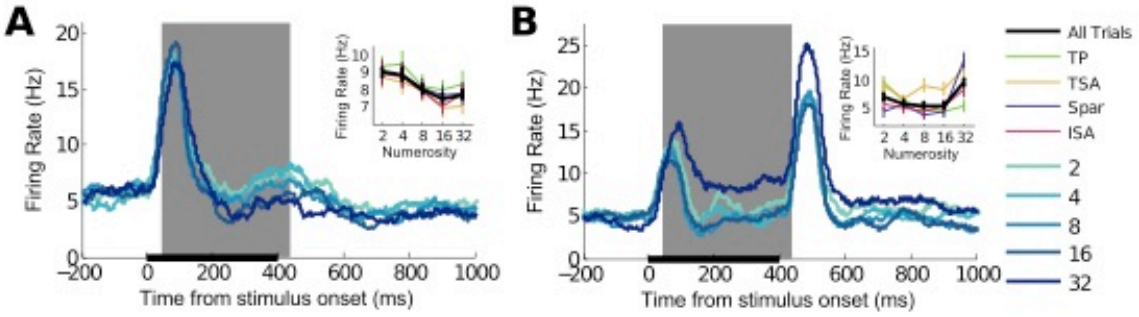

Figure 4-1. Example neurons modulated by stimulus features.
(A and B) Peristimulus time histograms aligned to stimulus onset for example neurons 1 and 2 respectively. Blue lines show the average response of the neuron to the presentation of the five numerosities (boxcar smoothed with a 75 ms kernel). The thick black line along the x-axis indicates the period that the stimulus is displayed (0-400 ms) and the light gray rectangle shows the analysis epoch (40-440 ms). The thick black plot in the inset shows the average response in the analysis epoch to the five numerosities. The colored lines show the average response to the numerosities considering the stimulus subsets separately. The green plots are the total perimeter (TP) controlled subset, the orange the total surface area (TSA) controlled, the blue the sparsity (Spar) controlled, and the red the item surface area (ISA) and item perimeter controlled.

To control for the effect of non-numerical visual features on the neural activity, we divided the stimuli into four subsets that controlled for total perimeter, total surface area, item sparsity (inverse of density), and individual dot size. For each of these subsets, one stimulus feature was varied orthogonally to numerosity over five levels. Within any one stimulus subset, numerosity necessarily covaried with some uncontrolled visual features, but by employing multiple subsets we were able to control for multiple features. As a result, a main effect of numerosity on firing rate, as observed in both of the example neurons in Figure 4-1 is a necessary but not sufficient condition defining a numerosity selective neuron.

We took two approaches to determine if each neuron was significantly modulated by numerosity. First we ran an ANOVA with factors for numerosity, stimulus subset, and an interaction term (Nieder and Miller 2004a; Roitman, Brannon, and Platt 2007; Viswanathan and Nieder 2013, 20). If a neuron is modulated by numerosity only, then there should be an effect of numerosity, but no effect of stimulus subset and no interaction. We found this to be the case for the first example neuron (Figure 4-1A inset; numerosity $p << 0.001$; stim. subset $p = 0.110$; interaction $p > 0.999$), but not for the second (Figure 4-1B inset; numerosity $p << 0.001$; stim. subset $p << 0.001$; interaction $p << 0.001$). The difference between the example neurons is apparent in the colored plot in the insets in Figure 4-1, where the effect of numerosity on firing rate is plotted separately for the different stimulus subsets.

Across the population of 118 VIP neurons, 12/118 (10.2%) neurons were modulated by numerosity. However only 5 of those neurons (4.2%) were significantly modulated by numerosity alone and thus showed no main effect of stimulus subset or interaction of numerosity and stimulus subset (numerosity $p < 0.01$; stim. subset $p > 0.01$; interaction $p > 0.01$). In contrast, 34/118 (28.8%) neurons were significantly modulated by one of the non-numerical stimulus parameters. However, using the approach outlined above and taken by previous researchers, we cannot say which non-numerical stimulus parameters were affecting number. Thus, to compare the encoding of numerosity with the encoding of other visual features on a single neuron level and gain a better

understanding of the stimulus information content of VIP on a population level, we need to develop a new analysis strategy.

In previous work, we put forth a novel mathematical description of dot array stimuli by algebraically demonstrating that eleven different visual stimulus features can be fully specified as the linear combination of three fully independent features; the numerosity, *Size,* and *Spacing* of the items within an array (DeWind et al., *under revision*). We use the terms *Size* and *Spacing* to denote these novel mathematical constructs. Intuitively, *Size* is the feature that changes when a fixed number of dots with fixed distance between their centers change in diameter. *Spacing* is the feature that changes when a fixed number of dots of fixed diameter are spaced further apart or are placed more closely such that the entire array occupies varying amounts of the visual field. On a logarithmic scale, these three features define eight other visual features via strictly linear equations: total surface area of the dots, individual surface area of each dot, the total perimeter or contour distance of all dots, the individual perimeter of each dot, the area of visual field subtended by the entire stimulus, the sparsity of the dots (visual field area per dot), coverage (proportion of the entire stimulus array occupied by dots), and apparent closeness (overall scaling of the entire stimulus). These features are defined mathematically in Appendix A.

Number, *Size*, and *Spacing* can be imagined as the three cardinal axes of a three-dimensional “stimulus space” illustrated in Figure 4-2 with the other eight features increasing and decreasing along different directions in the space.

Individual stimuli occupy a single point defined by the three-value vector of number, *Size*, and *Spacing*. These three values are sufficient to determine the other eight stimulus features algebraically. Stimuli occupying two different positions within this space define a vector and differ in a given feature to a degree proportional to the vector component along the dimension associated with that feature.

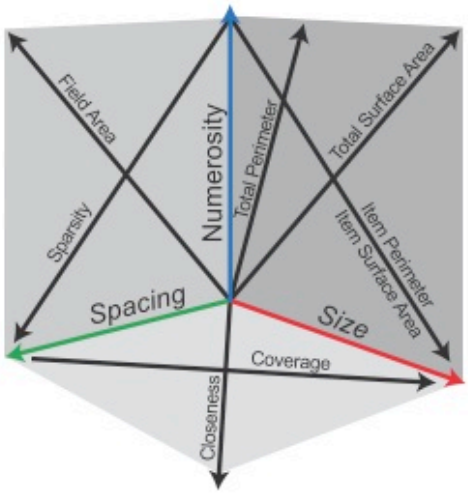

Figure 4-2. Three-dimensional stimulus space.
Stimulus arrays utilized in this and most other investigations of numerosity can be parsimoniously described using three values: numerosity, *Size*, and *Spacing*. These values can be thought of as coordinates in a three-dimensional stimulus space. Eight other stimulus features are defined by strictly linear equations when all features are on a log scale, and so can be visualized as lines (or rotated axes) within the stimulus space. Stimuli that differ in their position along a feature line differ in that feature proportionally to their distance along that line, whereas stimuli that lie on a plane orthogonal to that line all have the same value for that feature. The geometric constraints of this space provide an intuitive tool for visualizing the necessary collinearities between the visual features. Arrowheads indicate the direction in which that feature increases.

The three fully independent stimulus features are ideal regressors in a generalized linear model predicting firing rate. Given that the eight other features are linear combinations of number, *Size*, and *Spacing,* we can derive the effect of any of the eight other stimulus features on firing rate from these three regressors. This approach is superior to previous attempts to control for non-numerical

stimulus features because it treats numerosity simply as another visual feature and compares its ability to explain firing rate variance with the ten other visual features in an unbiased manner. As a result, we can directly test the hypothesis that VIP neurons are biased towards the representation of number.

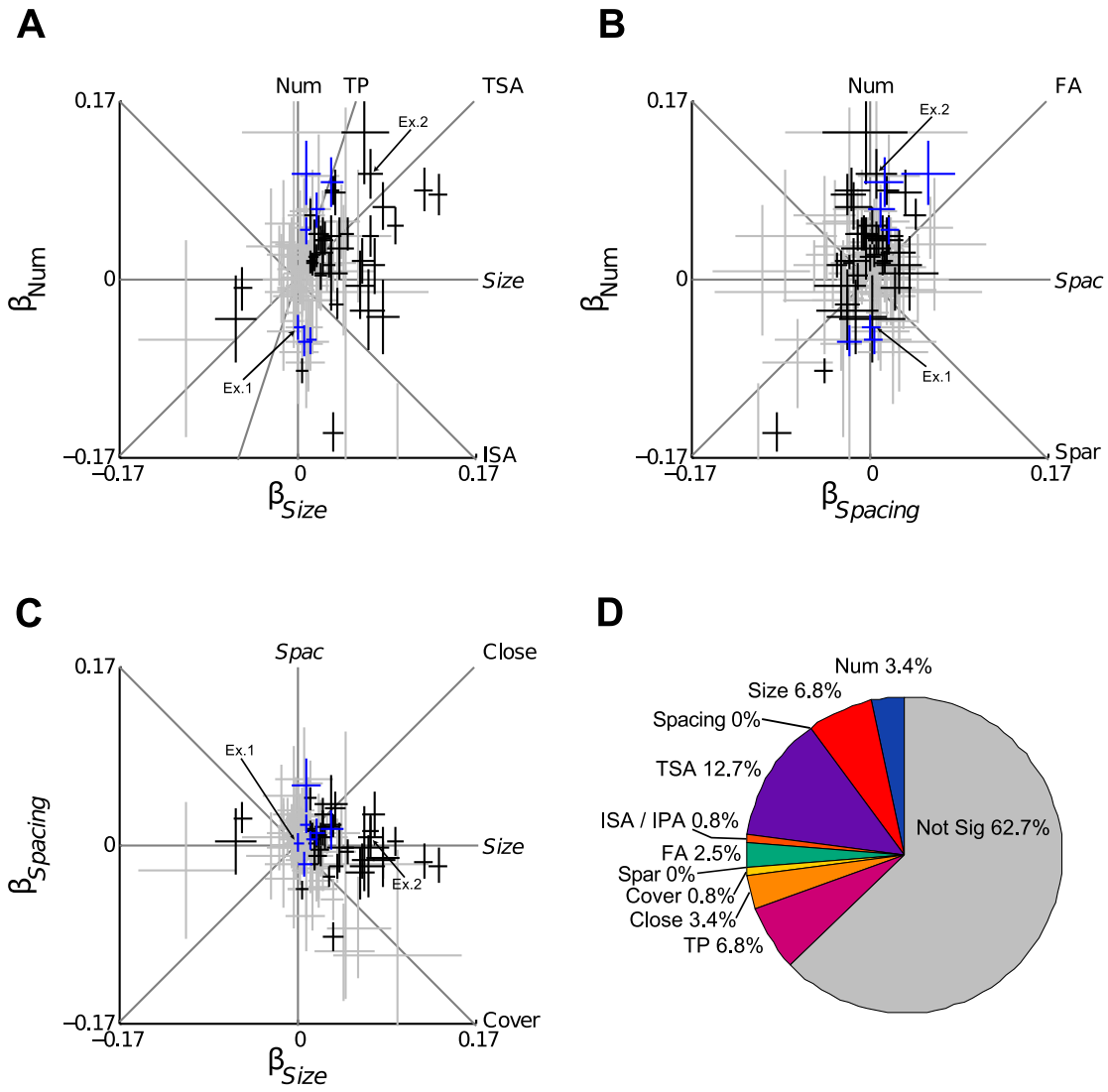

Figure 4-3. A heterogeneous population of VIP neurons are modulated by many visual stimulus features.

(A – C) Coefficients from the generalized linear model plotted against each other. Within a subplot, each cross represents the coefficients of a particular neuron and the size of the cross indicates standard error of the estimate. All neurons are plotted once in each subplot. Light gray crosses are not significant for the whole model ($p > 0.01$); blue crosses are neurons identified as "numerosity-only" meaning the numerosity coefficient is significant ($p < 0.01$) and the *Size* and

Spacing coefficients are not significant ($p > 0.01$); black crosses are model significant but not numerosity-only. The darker gray lines indicate where the coeffecients of hypothetical neurons encoding a particular feature fall (feature indicated by abbreviation). Example neurons from Figure 4-1 are indicated by Ex. 1 & Ex. 2 (D) Breakdown of the feature that best predicts changes in neuronal firing rate (based on reduced model AIC comparison). Neurons that were not whole model significant are categorized as "Not Sig". NB: This is an exclusive categorization that counts each neuron once. Some neurons categorized as "numerosity-only" were actually better explained by other stimulus features as a result of marginal effects of *Size* or *Spacing*.

Abbreviations: numerosity (Num), *Size* (*Size*), *Spacing* (*Spac*), total perimeter (TP), total surface area (TSA), items surface area/ item perimeter (ISA), field area (FA), sparsity (Spar), apparent closeness (Close), and coverage (Cover).

We can use the model to detect the influence of a non-numerical feature on firing rate. The effect of a non-numerical feature is equivalent to an exact ratio of the number, *Size*, and *Spacing* coefficients returned by the model. For example, an effect of total surface area would appear as an equal effect of number and *Size*, and no effect of *Spacing*. Sparsity appears as an equal but opposite sign effect of number and *Spacing* and no effect of *Size*. An effect of number appears as an effect of number and no other term.

Figure 4-3A - 4-3C are plots of the three model coefficients against each other for all 118 neurons recorded in VIP. These plots provide a useful visual summary of the representational properties of the neurons. For example, a neuron modulated by only numerosity would fall along the numerosity line in both 4-3A and 4-3B and at the origin in 4-3C. The example neuron 1 from Figure 4-1A is indicated in all three plots and is such a numerosity-only neuron, defined as being significantly modulated by number ($p = 0.001$) but not *Size* ($p > 0.999$) or *Spacing* ($p = 0.992$), confirming the original ANOVA analysis. As a result of the relationships between coefficients outlined in the previous paragraph and detailed in Appendix A, neurons modulated by other features would fall along the

corresponding feature line. For example, if a neuron were modulated only by total surface area (or luminance), then it would be significantly modulated by both numerosity and *Size* in equal proportions, but not by *Spacing*. Such a neuron would fall along the total surface area line in Figure 4-3A, the numerosity line in Figure 4-3B, and the *Size* line in Figure 4-3C. The example neuron 2 from Figure 4-1B is not far from the criteria for a total surface area neuron (β_{Num}= 0.10, $p < 0.001$; β_{Size} = 0.07, $p << 0.001$; $\beta_{Spacing}$ = 0.01, $p = 0.98$). This example also highlights the importance of correctly interpreting a main effect of numerosity; it is necessary but not sufficient for classification as a numerosity-only neuron. These examples also highlight the main benefit of the modeling approach pioneered here: the ability to discriminate between the effects of different visual features.

We found that the generalized linear model significantly explained variance in firing rate in 44/118 (37.3%) of neurons during the stimulus epoch ($p < 0.01$). It is apparent from looking at the “cloud” of model coefficients in Figure 4-3A – 4-3C that VIP neurons are not primarily positioned around the numerosity axis, indicating that they are not modulated by numerosity more than by other visual features. To quantify this impression, we calculated the visual feature that best explained firing rate modulation for each neuron that had significant variance explained by the model ($p < 0.01$). Figure 4-3D shows the proportion of the population best explained by each feature (reduced model with the lowest AIC, see experimental procedures for details). A minority of the significant neurons are modulated by numerosity with surface area (also luminance) and

total perimeter (also the proportion of the image at maximum contrast) claiming a large share of the neurons.

Our model assumes a linear effect of log features on the log firing rate of neurons; however, previous research has shown there are neurons tuned to individual numerosities in VIP (e.g. Nieder and Miller 2004a). Thus, it could be argued that there are more neurons specialized for numerosity (or other features) that were not detected by our linear model because they are tuned to intermediate values and do not show a monotonic trend. To look for such tuned neurons we adapted a non-parametric version of our model. We treated number, *Size*, and *Spacing* as categorical variables with *Size* and *Spacing* broken into quintiles.

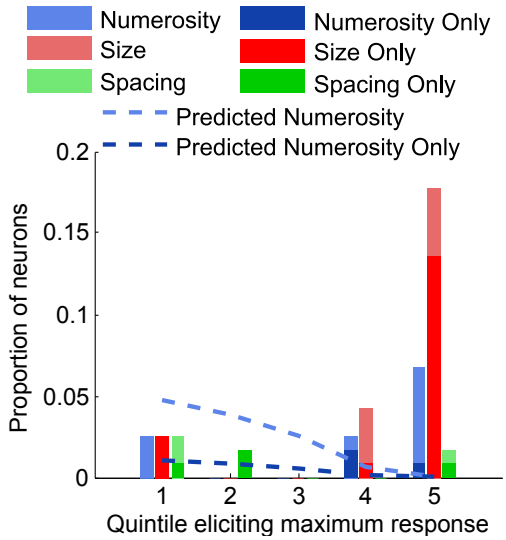

Figure 4-4. Non-parametric analysis reveals few neurons peaked at intermediate values. Bar plots show the proportion of neurons showing a main effect of numerosity, *Size*, or *Spacing* that fired maximally for each quintile (on a log scale). For numerosity quintiles were simply 2, 4 8, 16, and 32. Darker bars indicate the cells that were modulated by that feature ($p < 0.01$) and not by the other two ($p > 0.01$), whereas lighter bars were modulated by two or three features. The dashed lines are taken from the analysis of data from Viswanathan and Nieder (2013; see also Figure 4-S3). They represent the expected proportion of neurons firing maximally for each numerosity assuming tuned representations where the population of neurons tuned to a particular

numerosity decreases exponentially. The decay constant of this decay was fit to the previous data, and the total proportion of neurons has been scaled to match the total proportion found to be sensitive to the numerosity only neurons and the numerosity and any other feature neurons.

The results are summarized in Figure 4-4. Although many neurons were significantly modulated by at least one of the categorical stimulus features (38/ 118; 32.2%), we only found three numerosity-only neurons. We took this to mean that the monotonic assumptions built into the parametric model were valid on average. Furthermore, among the neurons that had significant variance explained by the model ($p < 0.01$) most fired maximally at one of the extreme values, indicating monotonic representations. Some cells did, however, fire maximally at intermediate values. Whether this is due to noisiness of cortical firing on top of a monotonic representation, as has been suggested (Qi Chen and Verguts 2013), or because the function relating firing rate to the stimulus is actually peaked is difficult to say, especially with so few numerosity selective neurons.

We tried to address this question for numerosity by extrapolating the proportion of neurons tuned to particular values observed by Viswanathan and Nieder (2013) to predict the number of neurons we could expect to see tuned to each of the five numerosities we presented. Viswanathan and Nieder (2013) only displayed the numbers 1-5, and we needed to extrapolate out to 32. We assumed that the distribution they observed was governed by an exponential decay function, whereby the number of neurons tuned to a particular number of dots dropped off as numerosity increased. We fit this function to the proportion of

numerosity neurons that they found to be tuned to 1-4, reasoning that all those neurons that appeared to be tuned to 5 were actually tuned to 5 and all values greater than 5. We fit the following exponential equation for λ:

$$T(n) = \lambda e^{-\lambda(n-1)} \quad \text{(Eq 4-1)}$$

T is the proportion of neurons tuned to numerosity n. We found the best fit lambda to the proportions reported by Viswanathan and Nieder was 0.28. Since we did not test all numbers between 1 and 32, we assumed that neurons tuned to numbers we did not test would fire maximally to the nearest value (on a log scale) that we did test. Finally, we scaled these proportions to match the small proportion of neurons we actually found that were sensitive to numerosity and were sensitive to numerosity only.

The dashed lines in Figure 4-4 show the result. In short, our findings are not consistent with the distributions observed by Viswanathan and Nieder under the assumption that they originated from actual tuned number neurons. In particular, they found that over 66% of the neurons they observed were tuned to the values 1-4. We would expect such neurons to fire maximally to the values 2 and 4 in our experiment. However, most of the numerosity-sensitive and numerosity-only neurons we observed were tuned to 16 and 32 (quintiles 4 and 5 in Figure 4-4). Thus, we think it is reasonable to conclude that the admittedly few numerosity neurons recorded here can be treated as monotonic over the range of values tested.

One of the key insights represented in our stimulus space and observed by others (DeWind et al., *under revision*; Dehaene et al., 2005) is that number is necessarily partially collinear with other visual features, even in ideally controlled stimulus sets. Studies focusing on numerosity typically treat non-numerical visual features as nuisance variables that need to be controlled in different trial subsets to rule out their effects on a dependent variable. However, there is another way of looking at necessary partial collinearity: across a broad set of stimuli, non-numerical stimulus features necessarily carry some information about number. Because we observed so few neurons that were modulated by numerosity, but a relatively larger population influenced by other features, we tested the idea that cells modulated by non-numerical stimulus features might contain information about numerosity across the population.

We constructed a classifier using the generalized linear model described above (parametric analysis) which could estimate the stimulus that generated a particular pattern of firing rates across the population. We generated each classifier trial by randomly picking a trial from each neuron during which they all saw a stimulus with the same set of parameters. Although these trials were all taken from different sessions, we treated them as pseudo-simultaneous to extract information from the entire recorded population. In cases where a given neuron was never presented with a stimulus of the given parameters, it was left out of that classifier trial. To avoid “double-dipping” our data set, we recalculated the model coefficients for each neuron, leaving out the trial being classified.

Using those coefficients for each neuron, we calculated the log-likelihood of each point in our stimulus space having caused the firing rate actually recorded on that trial. We normalized and averaged the log-likelihood functions of all the neurons together and picked out the maximum value in the space as the classifier's estimate. We repeated this procedure for 10,000 trials in each condition.

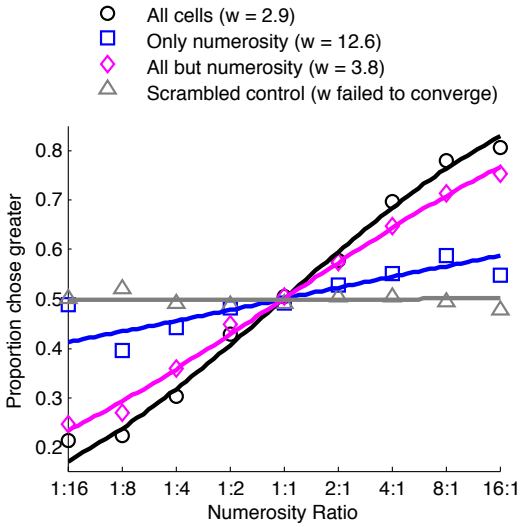

Figure 4-5. A stimulus classifier with access to the firing rate of non-numerosity-only neurons performs better than a classifier with access to numerosity-only neurons. Choices of classifier plotted against the actual ratio of the stimuli being compared. In the all cells condition the classifier had access to all neurons recorded. In the only numerosity condition the classifier only had access to the numerosity-only cells. In the all but numerosity the classifier had access to all cells but the numerosity-only cells. In the scrambled control estimates and the actual stimuli they derived from were scrambled. Lines represent the best fit psychometric function (Piazza et al. 2010).

On each trial the classifier made an estimate of the location of the stimulus in our three-dimensional stimulus space. To test the classifier we created an ordinal comparison "task" in which we randomly selected pairs of classifier trials and compared the numerosity estimates. Thus, we could get a binary "guess" of which stimulus the classifier viewed as larger. We plotted classifier pairwise estimates against the real numerical ratio of the pair of stimuli (Figure 4-5).

Interestingly, these estimates were well fit by the standard psychometric function allowing us to calculate a “neurometric *w*” or Weber fraction for the classifier. This *w* is analogous to the behavioral *w* used by Piazza et al. (2010) and many others. It is the standard deviation of the cumulative normal distribution describing the relationship between the frequency of choosing a stimulus and logarithm of the numerical ratio.

We first examined how well the classifier performed when it could draw information from the entire population of VIP neurons. Although the *w* is very poor by behavioral standards, the classifier performed far above chance, as assessed by the control using scrambled trial data. We then assessed whether the classifier would perform better using the small population of numerosity-only neurons, or using the larger group of non-numerosity neurons. We found that the classifier was able to better categorize the stimulus numerosity using the non-numerosity neurons than with the numerosity-only neurons.

The classifier estimated stimulus *Size* and *Spacing* in addition to numerosity, and so we could calculate an estimate for any stimulus feature using the equations in Appendix A. We could analogously run an ordinal comparison task for any feature and calculate a neurometric *w*. This results are shown in Figure 4-S4. In general, numerosity was a middling feature, not the largest nor the smallest *w*. In this sense the classifier performance mirrored the general finding from Figure 4-3D: that numerosity was not a stimulus feature that was

particularly well described, nor was it poorly described by the population response in VIP.

4.3 Discussion

4.3.1 A rich stimulus representation in VIP

We tested the hypothesis that the ventral intraparietal area is uniquely sensitive to the number of items in a dot array. Utilizing a unique analytic method, we tested which of eleven different visual stimulus features best explained neuronal firing rate. We found that there was a small subpopulation of neurons that were modulated by numerosity and no other feature, but that this population was equal or smaller in size and effect size to the subpopulations modulated by other features. We conclude that, while there are numerosity signals in VIP in animals naïve to numerical discrimination tasks, these signals are not particularly stronger or weaker than the signals encoding many other visual features.

Our findings instead suggest that VIP contains a rich representation of the visual features of a stimulus. All eleven stimulus features we examined could be discriminated to some degree by our classifier analysis. Our findings complement other single cell electrophysiology studies of VIP which have demonstrated complex visual response fields. VIP has been theorized to be at the locus of a system designed to represent peri-personal space (Graziano and Cooke 2006). These studies, however, have primarily focused on spatial and motion signals. Our research furthers our understanding of VIP, demonstrating that a rather rich

set of information about feature and identity is preserved in VIP, in addition to the spatial, motion, and numerosity signals observed previously.

Our findings are also in line with a separate set of research primarily based in humans that suggests that the human IPS is involved in comparison of magnitudes more broadly (for reviews see Bueti and Walsh, 2009; Cantlon et al., 2009; Hubbard et al., 2005; Walsh, 2003). All of the features we examined in VIP could be described as “visual magnitudes”. This finding is not unexpected in light of the results from Tudusciuc and Nieder (2007). They found that both line length and numerosity were encoded in IPS in partially overlapping and anatomically intermixed neuronal populations. Here we show that not only size and numerosity are encoded in VIP but many other visual magnitudes. These responses are present as purely visual responses with no incentive for the animal to attend to the stimulus and before any training to discriminate these dimensions. The IPS may not be unique with respect to number, but it may be an area uniquely suited to the representation and comparison of these types of magnitudes.

4.3.2 Tuned vs monotonic representations

A central feature of current models of numerical approximation is the tuned representation or labeled line code output layer of the putative neural network (Dehaene and Changeux 1993; Verguts and Fias 2004). In these models the tuned representations are considered the output layer of the neural network, and it is suggested that they form the neurobiological basis of the number sense and the ability to discriminate numerosities. Demonstrating that

individual neurons are tuned to particular numerosities requires demonstrating that some neurons are tuned to intermediate values between the minimum and maximum numerosities tested, because neurons tuned to the largest or smallest values tested cannot be differentiated from neurons related to numerosity by a monotonic function. In monkeys trained on a delayed match to sample task, many experiments have demonstrated the presence of tuned neurons in the IPS and DLPFC (e.g. Diester and Nieder 2007; Nieder 2012; Nieder and Miller 2004a; Nieder et al. 2002, 2006). These findings have been seen as a vindication of the computational models, a point that is especially convincing since errors are correlated with reduction of firing of neurons tuned to the sample numerosity (Nieder and Dehaene 2009).

The number sense is innate and does not require cultural transmission or training (Feigenson, Dehaene, and Spelke 2004). If neurons tuned to particular numerosities serve as the neurobiological basis of the number sense, then we should be able to observe them in monkeys naïve to numerical discrimination experiments. However, there is conflicting evidence for individual neurons tuned to individual numerosities in number-naïve animals. Roitman et al. (2007) found only monotonic encoding of numerosity in LIP in animals that were not trained to discriminate numerosity. In contrast, Viswanathan and Nieder (2013) found some neurons tuned to intermediate values in number-naïve monkeys. Even in that study, however, the majority of neurons recorded in both IPS and DLPFC were monotonic (or tuned to numerosities outside the range tested), with relatively few

tuned to intermediate values. A subsequent analysis by Chen and Verguts (2013) demonstrated that such a pattern of intermediate tuning might be an illusion generated by cortical noise while recording from an exclusively monotonic population. For example, a neuron apparently tuned to the second greatest numerosity tested may actually increase monotonically with numerosity, but have a slightly higher mean firing rate to the smaller numerosity due to chance resulting from noisiness in firing rates.

To determine whether there was a population of neurons tuned to intermediate numerosities or intermediate levels of any other feature, we ran a non-parametric version of our stimulus space analysis on each neuron. This analysis revealed only three numerosity-only neurons, making a quantitative analysis of intermediate tuning impossible. However, when we looked at tuning across the stimulus space, including neurons modulated by other visual features, we saw that most neurons were modulated by the minimum or maximum values tested for that feature, consistent with monotonic coding.

There may be neurons that are tuned to intermediate numerosities, and our analysis certainly cannot rule that out, especially given the finite ranges tested. However, our classifier analyses utilized our parametric model, and so monotonicity was a built in assumption. The classifier was able to discriminate numerosities successfully and replicated the behavioral finding of normal distributions of numerosity estimates. We therefore can conclude that even if tuned representations do exist, the decoder can successfully read out numerosity

by assuming monotonicity. In terms of the brain, it is possible that the decision processes that utilize numerosity information operate as if the code for numerosity is monotonic.

4.3.3 A distributed versus a sparse numerosity representation

The question of whether numerosity representations in IPS are tuned or monotonic is merely one aspect of a deeper question about the types of representations utilized by the nervous system. When numerosity is represented by individual neurons tuned to particular numerosities, the code for number is very sparse. Viswanathan and Nieder (2013) found that only 10% of VIP neurons were numerosity-only, and of those, less than 30% were tuned to values 5 or greater; this means that less than 3% of VIP neurons are responsible for the representation of all values greater than 4. If we accept the assumption that the number of neurons tuned to specific values follows the exponential decay function we fit to the Viswanathan data out to 32, we find that only 0.02% of VIP neurons represent that value and above (we found 2/118 or 1.7%); if we go further, only 0.000002% represent 64 and above. Of course, the details of our exponential function or fit could be off. However, the general point remains that if only 3% of VIP neurons represent all numerosities greater than 5, the code must be very sparse.

Are so few neurons the basis of our subjective sense of numerical magnitude? This is not a rhetorical question. The idea of extremely sparse representations was once mocked as the “grandmother cell” hypothesis (Rose

1996). However, more recently, very sparse representations have been identified (Hahnloser, Kozhevnikov, and Fee 2002; Quiroga et al. 2005; Vinje and Gallant 2000). If, however, the code for numerosity is monotonic, it is also a less sparse and more distributed code. Every monotonic number-only neuron could contribute to the subjective sense of number, not just those few that are tuned to it. A distributed representation might be more robust to neuronal cell death and cortical noise.

The results of our classifier analysis extend the concept of a distributed code for numerosity by demonstrating that there is more numerosity information outside the numerosity-only population than within it. A read out of numerosity that polls a distributed population level representation is better than one that relies on the numerosity-only neurons. Furthermore, not just numerosity but all features tested could be read out to a greater or lesser extent from this population code.

How can numerosity be read out from non-numerosity neurons? This depends largely on the insights of our stimulus space and regression model. If we assume features are represented on a log-compressed scale within VIP, then numerosity is a simple linear weighting of other features. For example, log-numerosity is equal to log-total-surface-area minus log-item-surface-area. This type of linear combination is trivial for the simplest feed forward neural networks. In general, a similar normalization algorithm is possible as long as there are

neurons that are modulated by cumulative features of the entire set as well as neurons that are modulated by features of individual items in a set.

According to this distributed view of feature encoding, the existence of numerosity-only neurons is partially epiphenomenal. Although they add to the amount of information about number and other features, they do not add unique information, and they are not the basis of the number sense per se. Numerosity and presumably other featural representations may be generated in untrained animals simply because they are useful for a generalist reconstruction of the visual stimulus. This idea is supported by the simulation study by Stoianov and Zorzi (2012), which demonstrated monotonic numerosity and total surface area virtual neurons after training a neural network using unsupervised learning. Critically, the learning algorithm did not train numerosity discrimination; instead, it trained the network to reproduce the visual stimulus as precisely as possible. Similarly, VIP contains a rich and robust stimulus representation from which many features can be reconstructed. This representation includes numerosity-only neurons, but perhaps only insofar as they contribute to a holistic stimulus representation.

4.3.4 Plasticity and training effects

Here we have suggested that multiple stimulus features are encoded monotonically by a heterogeneous population of neurons. Chen and Verguts (2013) have demonstrated that Viswanathan's and Nieder's data is consistent with a monotonic encoding scheme as well. What is the explanation of the tuned

representations of numerosity found by Nieder and colleagues in DLPFC and throughout IPS (both LIP and VIP) in many experiments (e.g. Diester and Nieder 2007; Nieder 2012; Nieder and Miller 2004a; Nieder et al. 2002, 2006)? We believe tuned numerosity representations in monkeys are the result of plasticity in response to training on a delayed match to sample task (DMTS), an idea suggested previously (Freedman and Assad 2009).

Categorical representations have long been known to exist in PFC after DMTS training. However, as sharp as category boundaries may be, when parametric stimuli are used, perceptual gray areas necessarily translate into degraded category representations. For example, merging cat-dog representations show attenuated activity in boundary conditions (Freedman et al. 2002). Similar categorical representations have been demonstrated in LIP and MIP (Freedman and Assad 2006; Swaminathan, Masse, and Freedman 2013). Although these studies focus primarily on the sharpness of the firing rate differences across boundaries, it should be noted that, near the boundaries, some attenuation of neuronal category preference is observed.

DMTS training for numerosities may similarly induce neurons to tune themselves to individual numerosities as if these numerosities were categories. Indeed, they must be treated as such in order to solve a DMTS task. A particular numerosity must be matched to itself, and near-misses are still misses requiring a sharply delineated categorical numerical concept. Other tasks, however, do not require such sharp boundaries between numerosities, and as a result may not

induce numerical representations that are so peaked. This possibility is supported by the Nieder ordinal comparison study, in which monkeys were trained to pick the stimulus with greater or fewer items based on a cue (Vallentin, Bongard, and Nieder 2012). In this task, rules were found to influence neural firing rate, but relatively few neurons were modulated by numerosity; neurons tuned to specific numerosities were largely absent.

Thus, numerosity signals are intrinsic to IPS before training as demonstrated previously by Viswanathan and Nieder and again here. However, the case for neurons tuned to individual numerosities in untrained monkeys is not so clear. It may be that tuned numerosity representations are the mark of particular types of training paradigms that favor individuated numerosity concepts. Such representations may actually be better understood as decision-making variables allowing the animals to maintain a template to compare to an upcoming test stimulus in the DMTS task, rather than veridical representations of the numerosity a visual stimulus.

A critical question is how these findings relate to the neural representation of number in humans. There is strong circumstantial evidence for tuned representations from functional imaging (Harvey et al. 2013; Jacob and Nieder 2009; Piazza et al. 2004; Piazza et al. 2007). Furthermore, humans think of and deal with numbers in categorical as well as parametric terms. However, human cultural and linguistic experience with numbers is very heterogeneous and thus

differs from DMTS training paradigms. Without recordings of numerosity neurons in humans, the nature of the neural code remains an open question.

4.3.5 Conclusion

We confirmed an earlier report that neurons in VIP are modulated by numerosity in monkeys before they are trained to discriminate numerosity (Viswanathan and Nieder 2013). We extended this finding by demonstrating that many other visual stimulus features are also encoded in VIP. We found that feature representations were largely monotonic. We also demonstrated that feature representations need not rely on the small number of neurons that vary in response to it and no other feature. Instead, representations that rely on many neurons perform better in psychophysical tasks. Our findings support a distributed and generalist stimulus representation in VIP that may be molded by task demands and attention to emphasize behaviorally relevant stimulus features.

4.4 Experimental Procedures

4.4.1 Subjects

Two adult male rhesus macaques served as subjects. Procedures were approved by the Duke University Institutional Animal Care and Use Committee.

4.4.2 Surgical and training procedures

A titanium head restraint prosthesis (Crist Instruments, Hagerstown, MD) and polysulfone recording chamber over the posterior parietal cortex (Crist Instruments) were surgically implanted under isoflurane inhalant anesthesia

using standard techniques. The animals received analgesics and antibiotics after all surgical procedures. The monkeys had been trained to perform fixation and visual orienting tasks for liquid rewards for previous experiments, but had never been trained on a task requiring the discrimination of numerosity or other magnitude. Neither had they been used in the implicit discrimination task previously implemented in the lab (Roitman, Brannon, and Platt 2007).

Eye position was monitored at 1000 Hz using the EyeLink system (SR Research, Ottawa, Ontario, Canada) during both training and recording sessions. All scripts were custom written in Matlab (Mathworks, Natick, MA) and stimuli were generated using Psychophysics Toolbox Version 3. Juice rewards were delivered via a tube placed in front of the monkey's mouth and controlled by solenoid valve.

4.4.3 Microelectrode and recording procedures

Before each recording session, the recording chamber was opened under aseptic conditions, cleaned using a 1% povidone-iodine solution, and repeatedly rinsed with sterile saline. A Teflon grid and X-Y micropositioner (Crist Instruments) were secured to the chamber and a Teflon coated microelectrode (FHC Inc., Bowdoin, ME) was lowered into VIP using a digital hydraulic microdrive (David Kopf Instruments, Tujunga, California).

Prior to initiating electrophysiological experiments, electrode trajectories into VIP were plotted using 0.5 mm slice magnetic resonance images taken in a 3T scanner at the Center for Advanced Magnetic Resonance Development at the

Duke University Medical Center. Each day, recordings were localized to VIP using stereotactic coordinates and by listening to characteristic sounds of gray and white matter while advancing the electrode. We focused recordings in the ventral-most portion of the intraparietal sulcus where multi-unit activity responded strongly to visual motion and where single cell firing was sensitive to motion direction in our response field mapping procedure (see below).

4.4.4 Neuron selection and response field mapping

Individual neurons were isolated in VIP and sorted based on waveform. VIP was identified based on stereotactic location, magnetic resonance images, the presence of a strong response to movement in the multi-unit response, and the presence of neurons with direction selectivity as determined by our receptive field mapping protocol. After the location of recordings was established, neurons were selected based solely on the quality of isolation.

The receptive fields of 91 of 118 neurons in this study were mapped systematically in two stages. First, motion sensitivity and direction selectivity was mapped: the entire display screen (44° visual angle) was filled with white dots (0.8**°)** on a dark background. The monkey fixated centrally for 200 ms, then the dots moved coherently in one of 8 directions (0**°**, 45**°**, etc. from straight up) at one of 3 speeds (40, 80 or 120 degrees/sec). Firing rate was averaged (75-150 ms window following motion onset sometimes shifted to capture peak activity), and an ANOVA was run with speed and direction as factors. Second, spatial sensitivity was mapped using a single dot (1.2**°**) that appeared briefly (133 – 400

ms depending on speed) in one of 25 positions on the screen defined by an evenly spaced 5 x 5 grid. In order to maximally drive cells, the single dot moved in the preferred direction at the preferred speed of the cell. In cases where the cell had no preferred direction, the dot moved in an arbitrary direction; in cases where the neuron had no preferred speed, the dot moved at 120° / sec. If the firing rate was significantly modulated by spatial location (ANOVA, $p < 0.05$, 70-170 ms window following motion onset sometimes shifted to capture peak activity) or, in the case of marginally significant ANOVA, if the spatial receptive field appeared to have a clear peak, then the cell was considered spatially selective, and the numerical stimulus was displayed in the location corresponding to the center of the receptive field. If the cell had no discernible spatial tuning, the numerical stimuli were displayed at the center of the monitor centered on the fixation point.

For the remaining 27 neurons, receptive fields were mapped informally by ear using a moving sinusoidal grating stimulus at different orientations, spatial frequencies, and locations within the visual field.

4.4.5 Numerical Stimuli

Stimulus parameters were constructed to control for both the intensive stimulus properties, which relate to the individual items within an array, and extensive properties, which relate to the entire stimulus. Intensive and extensive parameters cannot be simultaneously controlled with respect to number. For example, if individual item size (intensive property) is held constant, then

increasing numerosity will increase the total surface area of the stimulus as well as the total contour length or perimeter of the stimulus (extensive properties). Similarly, if the same amount of free space around each item (the sparsity or inverse density) is held constant, then the area of the stimulus array must increase with numerosity. By contrast, if the total surface area or contour length of the stimulus is held constant, then the individual items must become smaller as more are added; if the total array area is held constant, then each dot will have less free space, and sparsity will decrease (and density increase) as more items are added.

We constructed four sets of stimuli controlling for different intensive and extensive features of the stimuli within each set: the item size controlled set, the total surface area controlled set, the total perimeter controlled set, and the sparsity controlled set. In each set the numerosity of the array was 2, 4, 8, 16, or 32, with equal probability, and the features orthogonal to numerosity varied across numerosity. For example, in the individual item size set, individual item area and total array area varied orthogonally to numerosity; as a result, the total perimeter and total surface area increased with numerosity, and item sparsity decreased with numerosity.

4.4.6 Passive viewing procedure

Monkeys fixated (+/- 1.5 degrees) on a central rectangle to initiate the trial. Fixation was maintained for 1 s before a numerical stimulus randomly drawn from one of the four stimulus sets was displayed in the spatial receptive

field of the focal neuron for 400 ms. If the cell lacked a distinct spatial receptive field, the stimulus was centered on the fixation point (e.g. Nieder and Miller 2004a; Viswanathan and Nieder 2013). Fixation was maintained for 600 ms after stimulus offset. If the monkey maintained fixation for the entire 2 s period, he received a juice reward. Reinforcement did not depend on the numerosity of the stimulus.

4.4.7 Data analysis strategy: constructing independent dimensions for modeling

Following DeWind et al. (*under review;* Chapter 3), we described our stimuli using three independent parameters that together fully specify eleven stimulus features. These three parameters are the numerosity of the array and two novel terms, *Size* and *Spacing*. We use capitalization and italics to indicate that these are formally defined visual features, but we also wish to emphasize how these terms capture the intuitive concepts of item size and spacing. *Size* captures the single degree of freedom available in changing the size of a fixed number of items, which in turn affects the surface area of each item, the aggregate surface area of all items, the perimeter of each item, and the aggregate length of the perimeter of all the items. When numerosity is fixed, all of these other features are fully determined by *Size*. Similarly, *Spacing* captures the single degree of freedom available in changing how close or far apart a fixed number of items are. When numerosity is fixed, the field area, which is the area of the invisible circle within which the dots are drawn, and the sparsity, which is the field area per item, are all fully determined by the *Spacing*. Two other

stimulus features are independent of numerosity and depend only on *Size* and *Spacing*. One of these is apparent closeness, which denotes overall scaling of the stimulus without changes in relative proportions. The other is coverage, a term that denotes the proportion of the stimulus array field that is occupied by an item, regardless of the number of items or the overall scaling. Appendix A defines all of these features in terms of number, *Size*, and *Spacing*.

Another critical aspect of these eleven stimulus features is that, on a log-compressed scale, they are all related to numerosity, *Size*, and *Spacing* by linear equations (Appendix A). As a result, we can use numerosity, *Size*, and *Spacing*, as regressors in a generalized linear model of firing rate. From the coefficients returned by model fitting, we can also determine the effect of the eight other stimulus features. Different proportions of numerosity, *Size,* and *Spacing* effects correspond to effects of other features as determined by the linear equations relating the different features.

Spiking probability was assumed to be generated from a negative binomial distribution to account for over-dispersion compared to a Poisson. The generalized linear model we fit to each neuron is:

$$ln(FR) = \beta_0 + \beta_{Num} \log_2(Num) + \beta_{Size} \log_2(Size) + \beta_{Spacing} \log_2(Spacing) + \epsilon \quad \text{(Eq 4-2)}$$

4.4.8 Data analysis strategy: hypothesis testing

The three-value vector of β_{Num}, β_{Size}, and $\beta_{Spacing}$ define the best estimate of the feature changes that most modulate firing rate. This vector also defines a

line in the space in Figure 4-2. Lines in the three dimensional stimulus space define unique visual stimulus features, and so the neuronal β vector also defines the stimulus feature that modulates the neuron. For example, if β_{Num} and β_{Size} were significant and of the same magnitude and sign, it would be more parsimonious to say that the neuron was influenced by total surface area. The alternative is equivalent to saying that the neuron encoded item surface area when numerosity was fixed and numerosity when item surface area was fixed. On average and over the ranges tested, changes in the stimulus array along this neuronal feature vector dimension alter firing rate, and changes in the orthogonal plane do not.

We analyzed neurons that were whole model significant at $p < 0.01$, which is equivalent to the neurons that had a significantly non-zero neuronal feature vector. First we tested whether each neuron was numerosity-only, meaning that β_{Num} was significantly non-zero ($p < 0.01$) and that β_{Size} and $\beta_{Spacing}$ were not significantly non-zero ($p > 0.01$).

Second, we wanted to determine which single feature out of the eleven described in Appendix A was closest to the neuronal feature vector. We could do this because all models testing the effect of a single stimulus feature were reduced models of our main full model (**Eq. 4-2**). Thus we fit reduced models for every feature in Appendix A and compared the model fits using AIC. The cell was defined as encoding the feature with the lowest model AIC. The reduced models took the following form:

$$\ln(FR) = \beta_0 + \beta_{Feature} \log_2(Feature) + \epsilon \quad \text{(Eq.4-3)}$$

Note that numerosity neurons defined by this analysis were slightly different from those defined as numerosity-only. If we had extrapolated the numerosity-only test to the non-numerical features, it would not necessarily provide an exhaustive and mutually exclusive categorization of each neuron. For this reason, we used the AIC analysis to define the population breakdown shown in Figure 4-3D. The two analyses define similar and mostly overlapping populations of numerosity neurons.

4.4.9 Data analysis strategy: Classifier

To calculate the classifier estimate for a single stimulus we used the following procedure. First we identified all the neurons that viewed a particular set of stimulus parameters. A new unique stimulus was generated for every trial, so we treated stimuli with the same number, *Size*, and *Spacing* as equivalent. Among these neurons we then chose those that met the criteria for inclusion in the classifier condition (numerosity-only neurons, all but numerosity-only neurons, or all neurons). For each included neuron we identified a single trial on which it saw the stimulus being classified (sometimes neurons saw more than one such stimulus). To avoid “double dipping”, we then refit the model **(Eq. 4-2)**, leaving out the trial being classified. Using these new model coefficients, we densely sampled our stimulus space calculating the log-likelihood of each set of stimulus parameters given the firing rate on the selected trial. We repeated this process for all neurons that met the classifier condition criteria and were exposed

to the stimulus, creating a series of log-likelihood “stimulus space maps”. We then normalized the log-likelihood maps and averaged them all together. Each was weighted equally, regardless of how well the model fit. We then picked the location on the average map with the highest mean normalized log-likelihood. This was the parameter estimate (estimated number, *Size*, and *Spacing*) for the classifier trial.

Pairwise trials for the ordinal comparison task were created by randomly picking two classifier estimates and comparing them. If the classifier estimate of numerosity was larger for the actually larger stimulus, the classifier was correct. If the classifier thought that the stimuli were equal, then the larger stimulus was assigned by coin flip.

4.5 Supplementary Figures

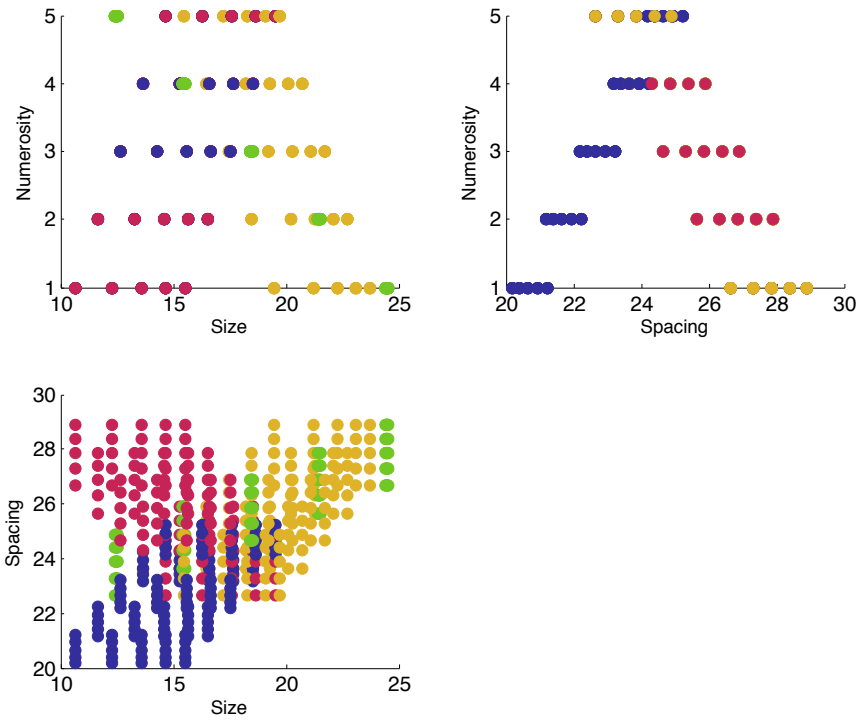

Figure 4-S1. All stimuli.
All stimuli used in the experiment with their numerosity, *Size*, and *Spacing* plotted against each other. Each dot is one stimulus (some dots overlap in some plots). Colors refer to subsets controlling for total perimeter (green), total surface area (yellow), sparsity (blue), and item surface area (red).

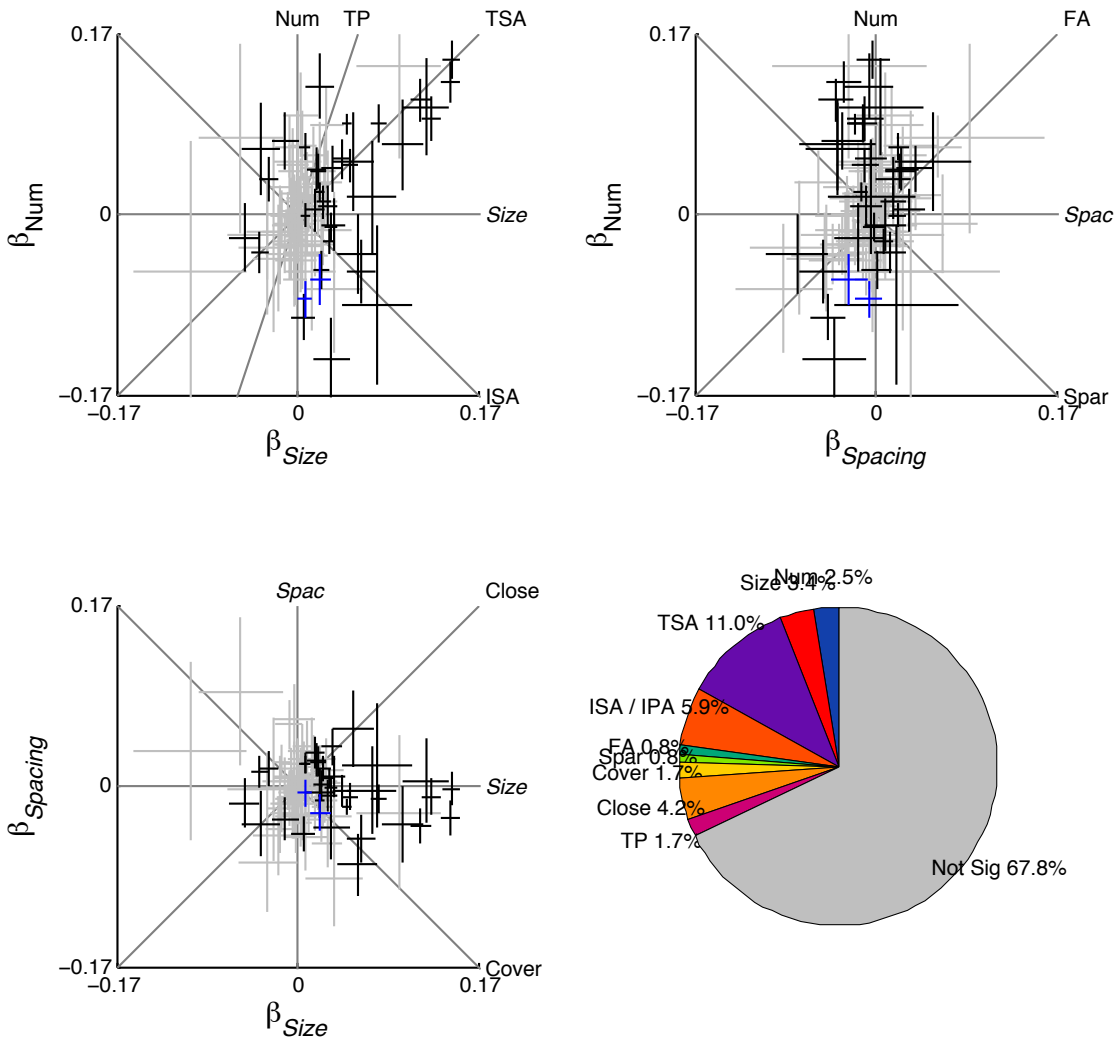

Figure 4-S2. Stimulus representation in the post-stimulus epoch.
The same plot and information as in Figure 4-3, but using the analysis epoch 40-440 ms after stimulus offset.

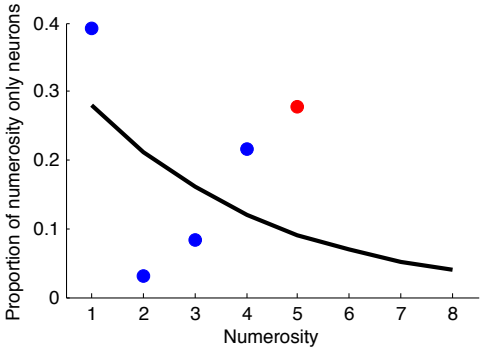

Figure 4-S3. Data from Viswanathan and Nieder (2013) fit to exponential decay function.
The proportion of numerosity only neurons that fired maximally for each numerosity tested (1-5). The black line shows the exponential decay function fit to the proportions for numerosities 1-4 (blue).

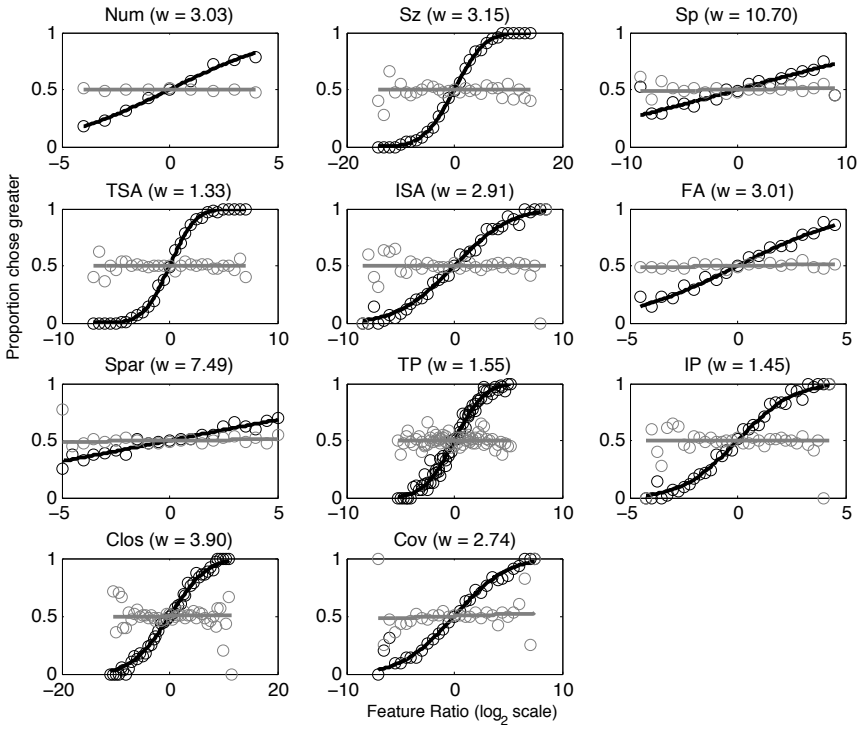

Figure 4-S4. Classifier discrimination for all features.
Black lines and markers show the all-neurons condition; gray show the scrambled control.

5. Evidence from pharmacological inactivation does not support a unique role for the intraparietal sulcus in approximate enumeration in macaque monkeys

5.1 Introduction

The “number sense” describes our intuitive ability to quantify sets of visual or auditory stimuli without counting them (Dehaene 1997). Educated adult humans share the number sense with infants (Izard et al. 2009; Xu and Spelke 2000), adults from cultures without verbal counting systems (e.g., Pica et al. 2004), and other vertebrates (Honig and Stewart 1989; Meck and Church 1983) including rhesus monkeys (Brannon and Terrace 1998). As a result, the number sense is described as a “core system of numerical representation” because it does not depend on language or cultural transmission (Feigenson, Dehaene, and Spelke 2004). It is theorized that the number sense may be combined with other core cognitive abilities to allow the development of the more sophisticated concept of exact number via education and development (Carey 2011), an idea supported by the finding that the acuity of approximate number representation is correlated with mathematical achievement in children (Qixuan Chen and Li 2014; Halberda, Mazzocco, and Feigenson 2008; Starr, Libertus, and Brannon 2013).

Much progress has been made towards describing the neural networks that subserve approximate enumeration. Functional imaging techniques have revealed that the horizontal segment of the intraparietal sulcus (hIPS) is activated during mathematical tasks (Dehaene and Cohen 1997), and that this activation is even more prominent during approximate calculation (Dehaene et al. 1999). Activity in hIPS varies with the number of items presented in a visual array even when participants are passively viewing the stimuli (Harvey et al. 2013; Jacob and Nieder 2009; Piazza et al. 2004). hIPS also responds to number words and written numerals (Piazza et al. 2007; Eger et al. 2003). As a result of these studies it has been theorized that the IPS subserves the representation of all approximate numerical magnitudes, and these representations can be activated by semantic stimuli or visual arrays of items (Nieder and Dehaene 2009).

Single neuron recordings in macaque monkeys have lent further strength to the idea that the IPS plays a critical role in numerical cognition. In monkeys performing a delayed-match-to-numerical-sample task, neurons in the IPS and the dorsolateral prefrontal cortex (DLPFC) contain single neurons that modulate their firing rate in response to the number of items in the sample stimulus (Nieder and Miller 2004a). Interestingly, the neurons in the IPS encode numerical information earlier than neurons in the DLPFC, indicating that number may first be extracted in the parietal cortex, and then relayed to prefrontal cortex where further processing allows numerical information to be used in decision making processes.

The IPS is not a single homogeneous region of cortex; multiple functionally and anatomically defined subregions have been identified. Nieder and Miller (2004a), found number selective neurons throughout the IPS. The greatest concentration of number neurons (~20%) was in the ventral fundus of the IPS, the ventral intraparietal area (VIP). Subsequent studies have confirmed the presence of number cells in VIP in both trained (Nieder 2012; Tudusciuc and Nieder 2007) and untrained animals (Viswanathan and Nieder 2013).

Roitman et al. (2007) recorded from the lateral bank of the intraparietal sulcus (area LIP) which is superior and lateral to VIP and found an even larger percentages of number neurons (~50%), perhaps because the stimuli were displayed in the predetermined visual receptive field of the focal neuron. Thus, in monkeys as well as in humans, the IPS is thought to play a critical role in numerical perception.

The neurons recorded in VIP and LIP differ in their response to number. The VIP neurons recorded by Nieder and Miller (2004a) were tuned to individual numbers of dots. For example, a “4” neuron fired maximally for 4 dots and progressively less for numbers further from 4. In this sense the VIP neurons were “tuned” to particular numbers. In contrast, LIP neurons recorded by Roitman et al. (2007) were split into two populations, one that increased firing with number and one that decreased firing with number. Both of these response patterns are “monotonic”. Interestingly, both tuned and monotnoic response profiles were predicted by previous modeling (Dehaene and Changeux 1993; Verguts and Fias

2004). The monotonic neurons are thought to represent an earlier processing stage, whereas the tuned neurons are the basis of numerical perception. Both stages are critical, however, and we predicted that inactivation in either area would disrupt numerical cognition.

Beyond correlative data, some evidence suggests that the IPS in humans plays a causal role in numerical cognition. Naturally occurring lesions in the IPS result in specific mathematical deficits, a condition known as acalculia (for review see Cohen et al. 2009). In addition to supporting a dissociation of mathematical deficits from other cognitive impairments, the lesion literature also supports the dissociation of rote retrieval of mathematical facts and quantitative abilities within mathematics, with quantitative deficits being specifically associated with IPS lesions (Cipolotti, Butterworth, and Denes 1991; Dehaene and Cohen 1991; Dehaene and Cohen 1997; Delazer and Benke 1997). Children with dyscalculia, a developmental disorder of mathematical reasoning, have thinner cortical gray matter in IPS compared to non-dyscalculic control children (Isaacs et al. 2001; Rotzer et al. 2008). Functional imaging also shows that the IPS is less active in otherwise normal dyscalculic participants compared to healthy participants during calculation tasks (Rotzer et al. 2009).

Relatively few studies have performed controlled interventions in IPS and observed the effect on numerical cognition. Several studies have used transcranial magnetic stimulation (TMS) to transiently alter IPS function and examine the resulting effect on number processing (Cappelletti et al. 2007;

Cohen Kadosh et al. 2012; Dormal et al. 2008, 2012; for review see Sandrini and Rusconi 2009). Although the consensus of these studies is that TMS to IPS can influence numerical discriminations, there is less agreement on the lateralization of these effects and their specificity to numerical processing. For example, one study found that duration and numerical estimation could be dissociated (Dormal, Andres, and Pesenti 2008), but another found that line length and number estimation could not (Dormal, Andres, and Pesenti 2012b). Direct current stimulation of the parietal cortex was found to improve or impair numerical abilities depending on electrode polarity (Cohen Kadosh et al. 2010).

One recent study examined the effect of pharmacological inactivation of parietal area 5 on a numerical task in monkeys (Sawamura, Shima, and Tanji 2010). This area, on the rostra-medial bank of the IPS, was previously demonstrated to contain ordinal numerosity signals (Sawamura, Shima, and Tanji 2002). Monkeys were required to perform five hand movements of one type followed by five of another type to receive a reward. Muscimol impaired accuracy more for this task compared to a control task that required the same hand movements but without the need to keep track of numerical information. This type of production task, in which an animal is trained to produce a number of sequential movements, is fundamentally different than a perception task in which an animal is trained to discriminate stimuli based on numerosity. The link between these types of tasks is tentative, and it is not clear that they both rely on the same “number sense” construct. It is also unclear if the close anatomical

position of neurons representing number in these tasks is more than coincidence or whether these findings generalize to humans. This study, however, provides the most direct test and the strongest evidence to date for a causal role of posterior parietal cortex in numerical processing.

While the above review provides strong evidence that the IPS is involved in numerical processing, there is less evidence, especially in non-human animals, that the IPS is uniquely dedicated to number. The IPS is known to be involved in many types of stimulus processing and motor planning; indeed the monkey IPS has been primarily studied outside the context of numerical processing. LIP is known to be involved in covert visual attention, saccade planning, and saccade initiation with individual neurons tuned to particular locations within the visual field (Bisley and Goldberg 2003; Colby and Goldberg 1999; Colby, Duhamel, and Goldberg 1996). The critical property of LIP that emerges across multiple studies is that it alters its response to stimuli based on their behavioral relevance and expected value (Kiani and Shadlen 2009; Platt and Glimcher 1999; Shadlen and Newsome 2001). VIP is characterized by strong motion direction tuning, including cells tuned to complex motion stimuli such as optic flow fields (Bremmer, Duhamel, et al. 2002). It is also a multimodal region with tactile, vestibular, auditory, and visual responses (Bremmer, Klam, et al. 2002; Duhamel, Colby, and Goldberg 1998; Schlack 2005). This pattern of response properties has been theorized to comprise a representation of self-

motion during three-dimensional navigation (Bremmer, Klam, et al. 2002) or of approaching stimuli in peri-personal space (Graziano and Cooke 2006).

Here we investigate the causal role of both VIP and LIP in numerical discrimination in the macaque monkey. We reversibly inactivated large strips of both regions with intracranial injections of muscimol in two monkeys. We assessed performance on an approximate enumeration task and on intermixed trials of a color discrimination control task using similar stimuli and identical response mechanisms. We reasoned that, for two tasks so similar in their perceptual and response requirements, if there were any specificity for numerical processing whatsoever, performance on the numerosity task would be selectively disrupted by LIP and VIP inactivation. Contrary to expectation, VIP and LIP inactivation did impair performance, but equally so in both the enumeration and color tasks. These findings suggest that the IPS is not a unique source of numerical processing in monkeys. While they may support a more general role of IPS in perceptual discrimination, without an unimpaired control task, we cannot rule out more fundamental perceptual deficits.

5.2 Methods

5.2.1 General Procedures and Behavioral task

All procedures were approved by the Duke University Institutional Animal Care and Use Committee. Two male rhesus monkeys (*Macaca mulatta*; 13kg and 19kg) were each implanted with a titanium headpost and a recording chamber over parietal cortex (Crist Instruments). Post-surgical MRI scans were

used to determine prospective injection sites, which were confirmed by single cell and multi-unit electrophysiological recordings. Recordings were made using a tungsten microelectrode (FHC). Juice rewards were delivered via a tube placed in front of the monkey's mouth and automatically controlled by a solenoid valve. Eye gaze positions were measure with an EyeLink 1000 system (SR Research) for online task control. Task software was run under MATLAB (MathWorks Inc.) using Psychophysics Toolbox 3 (http://psychtoolbox.org). Stimuli were displayed on a color LCD monitor (1,280 × 800 or 1,280 x 1024 resolution, 45 - 55-cm viewing distance depending of behavioral rig).

5.2.2 Behavioral Task

We trained monkeys on two tasks, a color task and a number task. In both tasks, monkeys fixated on a central target to initiate a trial. After 700 ms of fixation (within 2.5°), two arrays of dots on gray circular backgrounds (diameter 2.7° – 9.6°) were simultaneously displayed in two of the four visual quadrants around the fixation point (eccentricity 8° - 12°). The stimulus arrays were displayed for 70 ms. After the dot arrays extinguished, monkeys were required to wait an additional 400 ms for the fixation cross to extinguish. This was the cue to make a saccade to the location where the "correct" stimulus had been (contingencies described for each task below). The monkeys were rewarded with ~0.15 ml juice for a correct response. Error responses were followed by an auditory tone, and the trial was repeated until the monkey responded correctly. These correction trials were excluded from analysis. Each behavioral session

consisted of at least 2,300 intermixed color and number trials, excluding correction trials.

One stimulus array was always displayed in the upper visual hemifield and the other was always in the lower hemifield. For monkey Sh, the upper and lower arrays were placed randomly on the right or left side. For monkey Br, however, both the upper and lower arrays were always placed together on either the left or the right. We made this change after preliminary analysis of Sh's data showed that many effects were restricted to the hemifield contralateral to the injection site.

5.2.3 Number and color stimuli

The only difference then between the number and color task were the stimuli and the contingencies for choosing the correct stimulus. In the number task, the dots within each stimulus array were fully color saturated. Within each array, some dots were red and some were another color, either green or blue. The correct array was the one with the larger number of red dots. Whatever the number of red dots in the correct array, the incorrect array always had that same number of dots of the other color (green or blue). For example, if the correct array had 11 red dots and 5 blue dots, then the incorrect array would have 11 blue dots and 5 red dots. The total number of dots in each array was always equal to the other and was 8, 16 or 32 depending on the trial. For 16 total dots, the number pairs were 9:7, 10:6, 11:5, and 12:4. For 32 total dots the number pairs were 24:8, 21:11, 18:14, 20:12 and. For 8 total dots, the number pairs were

6:2 and 5:3. We used different numerical ratios to modulate task difficulty for each monkey, and not all numerical pairs were shown to both. Dots ranged from 0.20° to 0.68° in diameter. For half of all trials, the total dot area (pixels) of each color within each array was equal, and so dot size negatively correlated with number. For the other half of the trials, each single dot size was randomly and independently determined, and so the total dot area of each stimulus was positively correlated with number.

The color stimulus array pairs were identical to the number stimuli except that all the dots within an array were the same color, but the color of the dots varied between arrays. The color was a combination of red and either green or blue. The correct stimulus was the one with greater saturation of red. The relative ratio of red and either green or blue determined the trial difficulty. This resulted in arrays that were either yellowish (when combined with green) or purplish (when combined with blue). The monkey had to select the “redder” yellow or purple.

5.2.4 Electrophysiological Recording

Single unit recordings sessions were conducted before drug infusion to identify LIP/ VIP locations. Recordings were conducted using single tungsten microelectrodes (FHC, Bowdoin, ME), a dura piecing guide tube (23 gauge), and a Kopf (David Kopf Instruments, Tujunga, CA) hydraulic microdrive system. Neuronal signals were amplified, digitized, and stored for offline spike sorting (Plexon, Dallas, TX) to verify the quality and stability of neuronal isolations.

We used single-cell and multi-unit activity to confirm the predicted location of gray and white matter boundaries obtained from MRI. We informally mapped the areas by ensuring that LIP showed saccade related activity and that VIP showed strong responses to moving stimuli.

5.2.5 Intracranial drug administration

Muscimol (Sigma-Aldrich Co. LLC, St. Louis, MO), a GABA-a agonist, was used to transiently unilaterally inactivate VIP or LIP. Muscimol was dissolved in saline to a concentration of 5 mg/ ml. Saline was used as vehicle control. For each muscimol session, three sites that were identified as either LIP or VIP by MRI and electrophysiological recording were chosen for injection. 2 μl muscimol solution or saline were infused to each spot at a speed no greater than 1 μl /min. Seven or eight injections were administered to each animal in each condition in each experiment. The injection needle (outer diameter: 160 μm; Hamilton Co., Reno, NV) was inserted through the guide tube and driven to the depth of 0.5 ~ 1.0 mm under the lateral bank of IPS surface where neuronal signals had been obtained during previous recording sessions. Injections were done one at a time, and the needle was fully withdrawn and reinserted in a different grid hole for each site. It took approximately one hour to complete all three injections. The task was run for a few minutes (approximately 100 trials) for performance adaptation either before or during the injection. The task was restarted immediately after the needle was retracted from the third injection and ran until at least 2,300 trials were completed (approximately 2.5 hours).

5.2.6 Data analysis

Accuracy and response time (RT) were measured by custom task scripts. RT was defined as the time from cue offset to the time that the monkeys had shifted their gaze outside the fixation area. Main effects and interactions of muscimol administration were tested using ANOVA. Post hoc tests were only implemented when justified by significant ANOVA interaction terms, and Tukey's range test was used to correct for multiple comparisons.

5.3 Results

We trained two rhesus monkeys to discriminate stimuli based on numerosity (Figure 5-1A). To test the hypothesis that the IPS plays a unique causal role in numerosity discrimination, we injected muscimol or control saline into the ventral and lateral intraparietal areas in two separate experiments. To test the uniqueness of the IPS contribution to numerical perception, we also trained the monkeys on a color control task in which they had to choose the stimulus with a redder hue (Figure 5-1B). These trials were intermixed with the numerosity discrimination trials within a session.

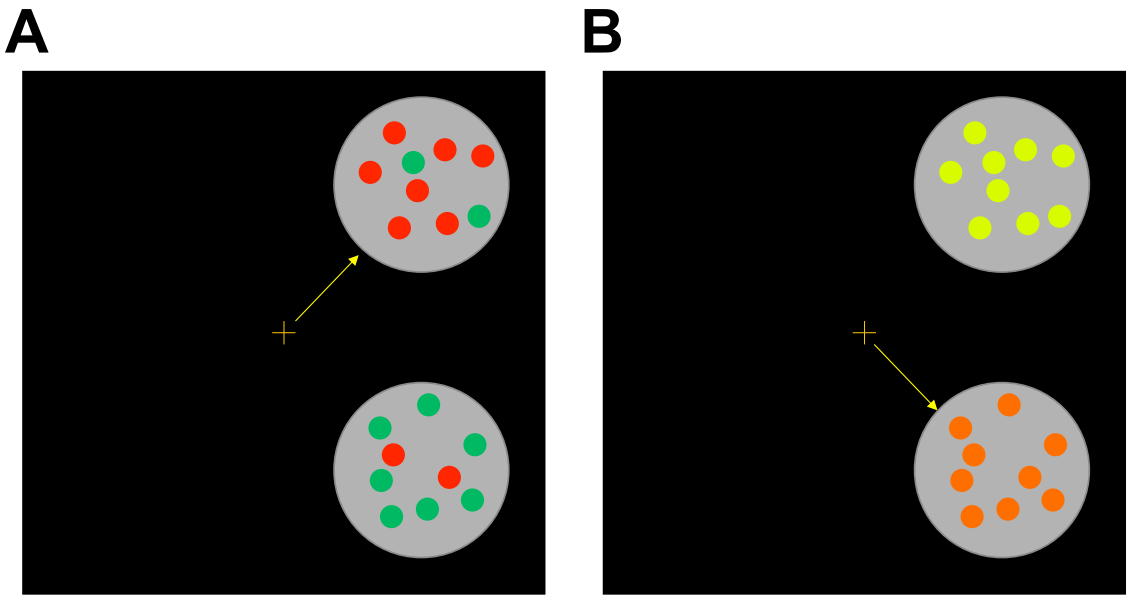

Figure 5-1. Number (A) and color task stimuli (B).
Monkeys fixated on a central stimulus. After 700 ms two stimuli appeared on the screen simultaneously, remained for 70 ms, and then disappeared. The fixation-cross remained extinguished after another 400 ms, after which the monkey could indicate a choice by saccading to the location where a stimulus had been. In the number task, the stimulus with the largest number of red dots was correct. In the color task, the stimulus with dots with more red hue saturation was correct. If the monkey made a correct response it was rewarded with juice; if it responded incorrectly the trial was repeated.

Both monkeys were able to perform the numerosity and color discrimination tasks well above chance expectations. The difficulty of the numerosity task was modulated by the numerical ratio of the red dots in the arrays being discriminated, allowing us to fit a standard psychophysical function to the performance and calculate the Weber fraction (*w*; Piazza et al. 2004). Similarly, performance on the color task was modulated by the difference in red color saturation between the stimuli.

5.3.1 Experiment 1: VIP inactivation

In Experiment 1 we unilaterally inactivated VIP in two monkeys and measured the effect on performance in the color and number discrimination

tasks. We ran a global ANOVA on accuracy and response time (RT) for each animal (see Appendix B for all ANOVA tables). The ANOVA contained four factors plus a full cross of all interaction terms. The first two factors were for drug (muscimol or saline) and task (number or color). The third was a term for the side on which the stimuli were presented (contralateral or ipsilateral to injection site). The fourth was for time within session (first half or second half); we reasoned that muscimol might take some time to diffuse into the tissue, or alternatively that the effects might begin to wear off. The critical test of our hypothesis would be revealed in interactions that included both drug and task. We were open to the possibility that these effects might only manifest themselves within the visual hemifield contralateral to the injection site, or might be most pronounced in either the first or second half of the session.

We found no main effect of drug on accuracy and no significant interactions that included drug and task in either monkey. We did, however, find a main effect of drug slowing response times in both monkeys. We also found a significant drug by task effect and drug by task by time effect on response time in monkey Sh. Post hoc tests showed that responses were more slowed by muscimol in the color task than in the number task, and that this difference grew from the first half to the second half of the session. We found no significant interactions that included drug and task in monkey Br. Figure 5-2 shows the results of the critical drug by task analysis in both monkeys.

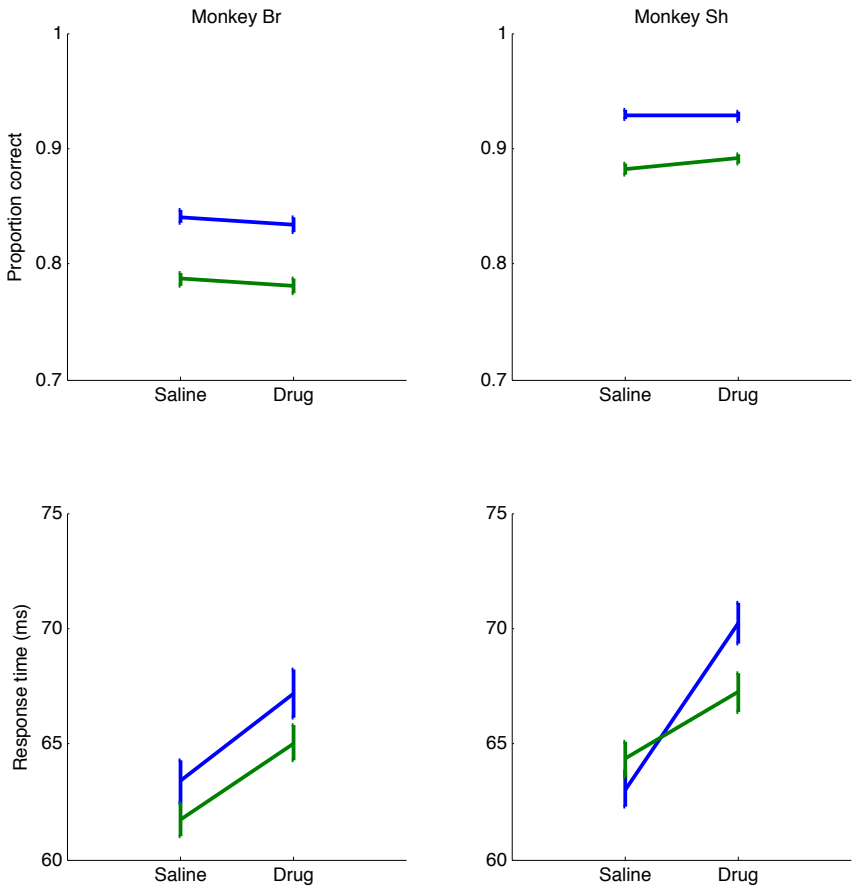

Figure 5-2. VIP muscimol impaired RTs, but no more in the number task than in the color task.
Accuracy (top panels) and RT (bottom panels) in Experiment 1 broken out by monkey (left and right columns) and task (blue line shows color task data and green line shows number task data).

We also observed a significant drug by side effect on response times in Br and a significant drug by side by time effect on response times in monkey Sh. Post hoc tests showed that responses were more delayed by muscimol in the visual hemifield contralateral to the injection site in both monkeys, although this effect did not emerge until the second half of the session in monkey Sh.

To investigate how the psychophysical characteristics of the task interacted with injected muscimol, we modeled the effect of ratio on numerical acuity (Piazza et al. 2004). This analysis revealed that there was no overall effect

of VIP muscimol on enumeration acuity in either animal. Similarly we found no effect of muscimol on the hue saturation effect in the color task (Figure 5-3).

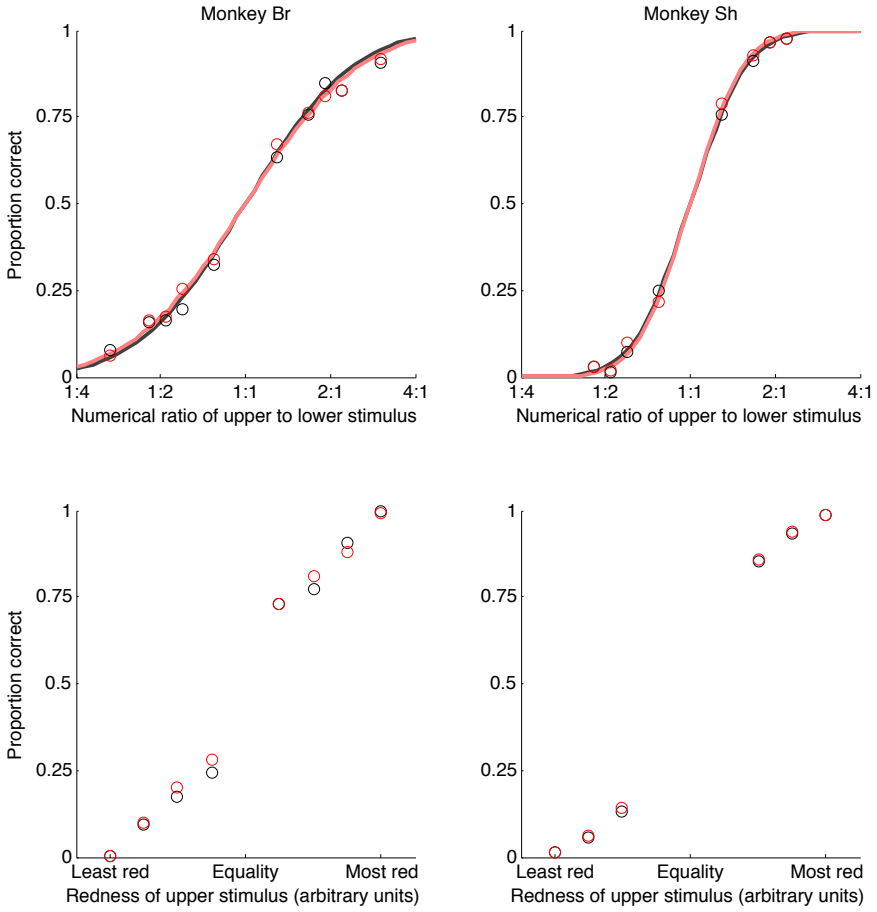

Figure 5-3. VIP muscimol has no discernable effect on number task or color task accuracy. Accuracy in the number tasks (top panels) and color task (bottom panels) for both monkeys (columns) broken out by drug condition (saline is black and muscimol is red). Psychophysical functions were fit to the number data (red and black fit lines). Numerical acuity (*w*) was not significantly affected by drug condition in either monkey (Table 1).

Table 5-1. Numerical acuity is unchanged after VIP muscimol.
w fit to muscimol and saline VIP injection data in each monkey

EXPERIMENT 1: VIP	**BR**	**SH**
W_{SALINE}	0.71 +/- 0.038	0.37 +/- 0.019
$W_{MUSCIMOL}$	0.75 +/- 0.042	0.35 +/- 0.017

5.3.2 Experiment 2: LIP inactivation

In Experiment 2 we unilaterally inactivated LIP in the same two monkeys and measured the effect on performance in the same color and number discrimination tasks. The data were subjected to the same global ANOVA on accuracy and response time (RT) for each animal. Figure 5-4 summarizes the same key results as Figure 5-2, but for LIP.

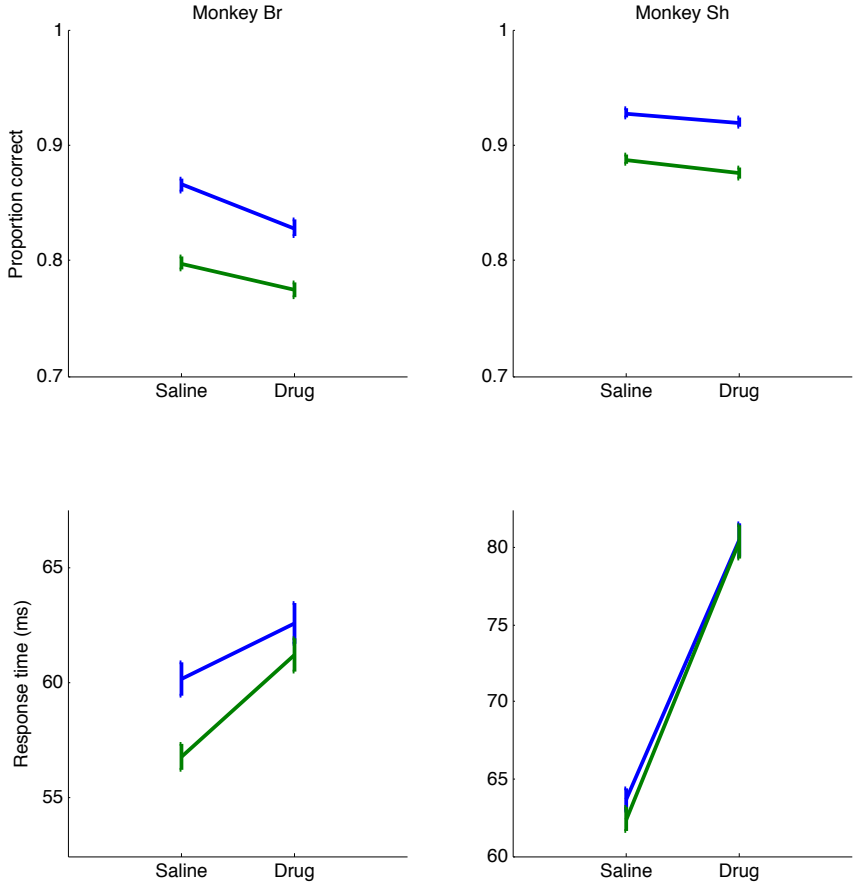

Figure 5-4. LIP muscimol impairs accuracy and RT in both monkeys; however, there was no evidence of selective impairment in the number task.
Accuracy (top panels) and RT (bottom panels) in Experiment 2 broken out by monkey (left and right columns) and task (blue line shows color task data and green line shows number task data).

There was a main effect of drug on accuracy for both monkeys. In monkey Br there were no significant interaction terms that included both drug and task. In monkey Sh, however, we found a marginally significant side by drug by task interaction. The post hoc test on Sh's data showed that muscimol impairment only occurred in the contralateral hemifield; within the contralateral hemifield only the color task was significantly impaired. Monkey Br showed a significant side by drug interaction and showed the same pattern as Sh except that both tasks were

equally impaired in the contralateral hemifield. Figure 5-5 shows the accuracy data broken out by drug, task and side for both animals.

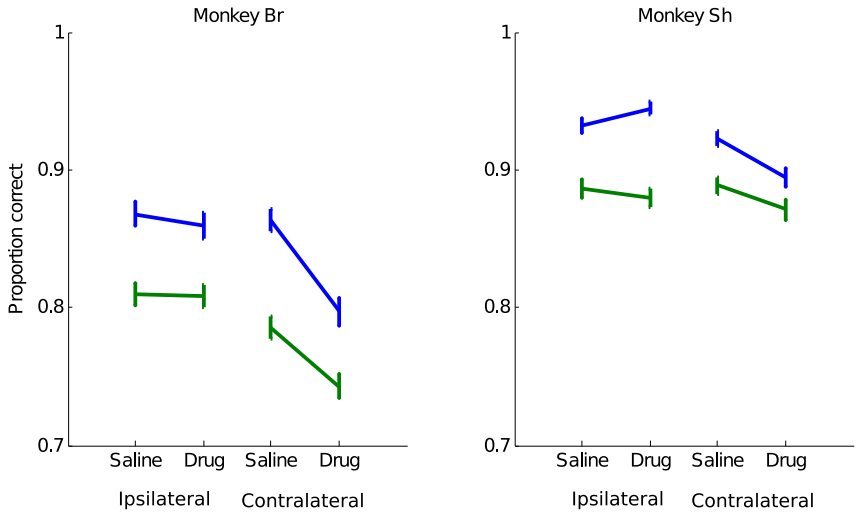

Figure 5-5. In the hemifield contralateral to LIP injection, accuracy in monkey Br was impaired for both tasks, but in monkey Sh only the color task was significantly impaired. Accuracy in Experiment 2 for each monkey (left and right columns), as a function of task (blue line shows color task data and green line shows number task data), and stimulus presentation side (plots either labeled Ipsilateral or Contralateral along the x-axis).

Muscimol to LIP significantly slowed RT in both monkeys, but did not significantly interact with drug or task in either monkey. Both monkeys had a significant side x drug interaction. Interestingly, when muscimol was injected, Br responded faster to stimuli ipsilateral to the injection site, and slower to stimuli contralateral to the injection site. Sh, however, responded more slowly to all stimuli under muscimol, although this effect was greater on the contralateral side. Both monkeys showed significant drug x time interactions, and post hoc tests showed that the drug increasingly delayed responses over the course of the session. The increase in delay was significantly greater in the contralateral hemifield (side x drug x time interaction) for monkey Sh.

We examined the ratio and hue saturation effects in more detail (Figure 5-6). We restricted this analysis to the contralateral hemifield, where the effect of LIP muscimol was stronger in both animals. We confirmed that accuracy was impaired by LIP muscimol in Br, which flattened the effects of numerical ratio and hue saturation in the number and color tasks respectively. In subject Sh, the hue saturation effect was attenuated, but there was no significant effect of muscimol administration on numerical acuity (*w*).

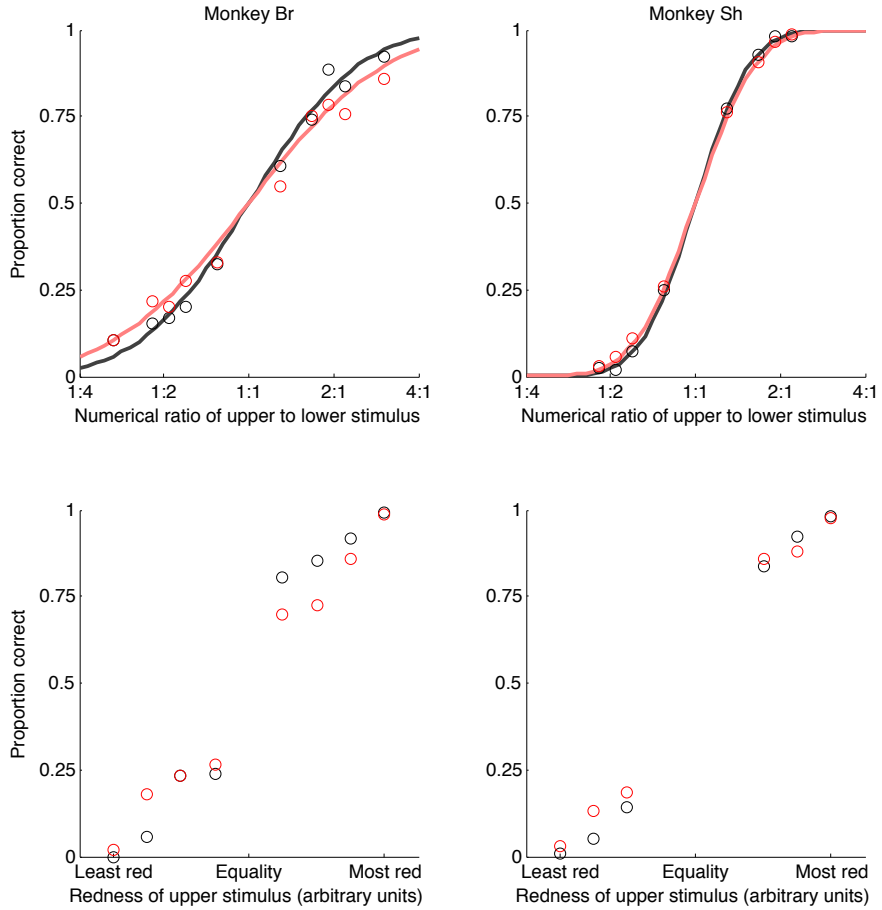

Figure 5-6. LIP muscimol impairs contralateral accuracy in both tasks in Br, but only in the color task in Sh.
Accuracy in the number tasks (top panels) and color task (bottom panels) for both monkeys (columns) broken out by drug condition (saline is black and muscimol is red). Psychophysical

functions were fit to the number data (red and black fit lines). Numerical acuity (*w*) was significantly affected by drug condition in Br but not Sh (Table 2).

Table 5-2. Numerical acuity is significantly decreased after LIP muscimol in monkey Br but not monkey Sh.
w fit to muscimol and saline LIP injection data in each monkey. Functions were fit only to data obtained in trials presented in the hemifield contralateral to the injection site.

EXPERIMENT 2: LIP	**BR**	**SH**
W_{SALINE}	0.72 +/- 0.054	0.36 +/- 0.024
$W_{MUSCIMOL}$	0.90 +/- 0.080	0.39 +/- 0.027

5.4 Discussion

We found deficits in RT after injection of muscimol to VIP, and deficits in RT and accuracy after injection of muscimol to LIP in two animals. However, these effects were no more evident in the number task than in the color control task. Thus, our data do not support the theory that LIP and VIP play a unique causal role in numerical perception and discrimination. How can this be reconciled with the large body of literature that shows the involvement of IPS in number processing in humans and monkeys?

One possibility is that, our results do not generalize to humans. Most of the human evidence for IPS specialization for number comes from fMRI studies, with no comparable studies in monkeys. For example, the Piazza et al. (2004) experiment habituated participants to a stimulus stream that fluctuated in all parameters except number, which was held constant. They then presented a dishabituation stimulus in which number varied as well. They used a whole brain analysis to determine which areas were sensitive to the change in visual number.

They found that the IPS was uniquely modulated by number when other features were controlled.

Single cell electrophysiology in monkeys has demonstrated that the IPS and PFC contain numerosity neurons. However, the method is spatially restricted. We do not know which areas do not have number neurons. This point is particularly important, because monkeys, unlike the humans used in imaging studies, usually have tens or hundreds of thousands of trials of experience distinguishing numbers. This type of training may influence the number of areas encoding number and the strength of that encoding. Without imaging studies of trained and untrained monkeys that parallel studies like Piazza et al. (2004) we cannot know which monkey brain areas are particularly specialized for numerical processing.

In addition to VIP and LIP, numerosity sensitive neurons have also been found in DLPFC. It has been theorized that IPS first extracts numerosity and then forwards the information to DLPFC, an idea supported by the observation that the IPS encodes number earlier than DLPFC (Nieder and Miller 2004a). However, another possibility is that DLPFC derives numerical information independently of IPS. It remains possible that if we inactivated IPS and DLPFC simultaneously, we would see a more dramatic numerical deficit. DLPFC, however, is involved in many learned categorical discriminations (Freedman et al. 2002; Kim and Shadlen 1999; Wallis, Anderson, and Miller 2001) and thus inactivation would likely impair performance for other discriminations.

Another possibility is that IPS activity is required for numerical processing, but that after unilateral inactivation, animals can compensate by relying on the IPS in the other hemisphere. We think this explanation is unlikely, however. If the unaffected contralateral IPS were fully compensating following the injection, we would not have seen any deficit in either task. However, we did find deficits in behavioral performance and these were greater in the contralateral visual hemifield in both experiments. If the IPS were specialized for number, then to the extent that we did see a deficit contralateral to the injection site, we would also expect that deficit to uniquely affect the number task. Instead, we found that both tasks were impaired, suggesting no specialization of IPS.

Another possibility is that the IPS is specialized for the perception and comparison of magnitudes such as time, space, or size, rather than only for numerosity (for reviews see Bueti and Walsh 2009; Cantlon, Platt, and Brannon 2009; Hubbard et al. 2005; Walsh 2003). Functional imaging studies in humans demonstrate that comparing different magnitudes (such as spatial location, luminance, and size) activates overlapping areas in the intraparietal sulcus (Cohen Kadosh and Henik 2006; Pinel et al. 2004; Zago et al. 2008). Single cell recordings in monkeys by Tudusciuc and Nieder (2007) have demonstrated that both numerosity and line length are encoded in IPS in partially overlapping populations. It is possible that our color control task was perceived is an ordinal magnitude comparison, with “reddishness” as the magnitude being compared. As a result, the IPS was engaged in both the numerical and color comparisons.

Another untested possibility is that any task that relies on a two item forced choice would be dependent on the IPS. LIP has been implicated in discerning the value of a saccade to a specific location in the visual field (Bisley and Goldberg 2003; Platt and Glimcher 1999; Shadlen and Newsome 2001). More recently it has been demonstrated that LIP encodes arbitrary boundaries in random-dot-motion direction when those boundaries predict reward contingencies and even when the distinctions are not associated with particular saccade targets (Freedman and Assad 2006; Freedman and Assad 2009). Similar neurons have also been found in the medial intraparietal area, and may be found throughout posterior parietal cortex (Swaminathan, Masse, and Freedman 2013). Thus it is possible that, in monkeys, the number neurons found in IPS are part of a more general system that encodes behaviorally relevant categorical distinctions in parametrically varying stimuli. This generalized system for perceptual decision-making may discriminate dot-motion, number, color, or any other perceptual feature that is tied to reward. Thus, when we inactivate IPS, we degrade an animal's ability to perform any such discriminations.

6. Conclusions

6.1. Research questions revisited

6.1.1 The question of malleability

In Chapter 2 we directly tested the malleability of the number sense in adult humans. We found that simple trial-by-trial feedback improved accuracy on an ordinal non-symbolic numerosity comparison task. Thus, we could conclude that the number sense changes with training and experience, even in adulthood.

However, we were left with the question, what was the mechanism of this change? In our data we found that both *w*, the metric of number sense acuity, and surface area effect index, the metric of bias, decreased with training. In the discussion in Chapter 2, we proposed decreasing reliance on stimulus surface area as a potential mechanism by which acuity increased (and *w* decreased). However, we also acknowledged that this could not be the entire story, because *w* dropped even for the subset of trials in which total surface area was held constant.

When I reflected on the results and interpretation of this experiment, I realized that the surface area effect index was insufficient to fully capture the effect of non-numerical stimulus features on the number sense. For example, when total surface area was held constant, item surface area was inversely related to numerical ratio. As a result, I could not be sure whether the increase in numerical acuity was due to decreasing reliance on non-numerical stimulus

features, or acuity itself. Indeed, on a fundamental theoretical level, I was not sure if these two concepts could be distinguished from each other.

In Chapter 4 we examined the malleability of the neural basis of the number sense. Previous studies had demonstrated neural “tuning” to individual numbers of items. A particular neuron might be tuned to “4” dots on the screen, firing maximally to four and less to three or five items. At the time we started the experiment, all previous studies of monkey VIP were made in monkeys performing a delayed-match-to-numerical-sample task. We wanted to know whether tuned number representations were endemic to the brain or whether they were the result training. To answer this question, we recorded from monkeys that were naïve to numerosity discrimination tasks.

We began the experiment using the standard method of ruling out the influence of non-numerical stimulus features: we created a list of features that might influence neural firing rate and then created subsets of stimuli in which those features did not vary with number. The nature of dot array stimuli meant that when one stimulus was controlled, others varied, but with enough control conditions we could rule out each feature in turn. To this end, we created four stimulus control conditions: total perimeter controlled, total surface area controlled, item surface area controlled, and sparsity (and density) controlled.

This allowed us to identify “numerosity-only” neurons: neurons modulated by number and no other stimulus feature. When we had finished data collection, we found that very few neurons met this criterion, especially compared to

previous reports (e.g. Nieder and Miller 2004a). This implied that the delayed-match-to-sample training did indeed affect the representation of number in VIP; in particular, training seemed to increase the number of neurons encoding numerosity.

With so few numerosity-only neurons, it was difficult to determine if the cells we did observe were tuned to individual numbers as observed in previous VIP recordings (e.g. Nieder and Miller 2004a) or varied monotonically with number as observed in the previous LIP recordings using monkeys with less numerical training (Roitman, Brannon, and Platt 2007). One piece of evidence against tuned neurons was the dearth of numerosity-only neurons (and numerosity-and-other-feature neurons) that fired maximally for intermediate values tested. Another piece of evidence against tuned representations came from comparing our distribution of peak firing rates to the distribution predicted by our re-analysis of the data from Viswanathan and Nieder (2013), which also recorded VIP neurons in untrained monkeys. They only used the numbers 1-5 in their study, while we used 2-32. By extrapolating from the 1-5 distribution, we found that less than 1% of VIP neurons would be expected to be tuned to all numbers greater than 32. The distribution did not fit our data set, and we found these predictions to be unbelievable. Rather, a cleaner interpretation of the data is that both the Viswanathan and Nieder data set and our data set were generated by neurons with firing rates that were monotonically related to stimulus number. The occasional neuron firing maximally for an intermediate value is

better explained by noisiness with limited sample size (Qi Chen and Verguts 2013).

6.1.2 The problem of stimulus control

I wanted to move beyond simply classifying neurons as either numerosity-only or not. I knew we could tell if non-numerical features modulated firing rate, but with the existing analysis techniques I could not tell which ones. At the same time, I knew we could do better. I had intentionally designed my stimuli such that non-numerical features were not just held constant while number varied, but also varied orthogonally to number. There had to be a way to leverage this variance. I entered a long conceptual detour, much aided by many conversations with lab mates. The goal was to treat number and other visual features equally, and to get an unbiased estimation of which feature or features did affect firing rate. The final result is the stimulus space concept described in Chapters 3 and 4.

The essential nature of the solution was to identify the degrees of freedom in the multitude of partially covarying visual features of a dot array stimulus. In particular, the dot array stimulus can be described with just three numbers. Although this description is not exhaustive, it is very rich; we identified eight other visual features that were fully determined by our three.

We used the logic of the stimulus space in a regression model. The model allowed us to correctly attribute variance in a dependent variable that resulted from changes in number or any other visual feature. We found that this solved the question raised by the experiment in Chapter 2: can numerical acuity (*w*) be

differentiated from biasing effects of non-numerical features? By applying our model to human behavioral data in Chapter 3, we answered that question both theoretically and empirically: yes. Acuity and bias are differentiable concepts.

6.1.3 The question of IPS specialization

The stimulus space regression approach also led to a breakthrough in the analysis of our neural data. Once we overcame the stimulus control problem, we were able to provide a rich description of VIP neural representation. We found that many visual features were simultaneously represented by VIP neurons. Numerosity, it turned out, was merely a “middling feature”, neither exceptionally over nor under represented in VIP.

This view, however, depends on an analysis of individual neurons. We also wanted to see if there was information about number at the population level and if this information could guide a simulated behavioral decision. Our classifier estimated stimulus parameters from the firing rates recorded across the population. We found that we could robustly estimate numerosity using this method. Although the performance was poor by behavioral standards, the standard psychophysical functions (Piazza et al. 2010) fit the data very well, indicating that some similar mechanism might actually be guiding behavior.

Interestingly, we found that classifier performance did not depend on the numerosity-only neurons. In fact, removing these neurons from the classifier’s pool only marginally disrupted performance. Removing all the other neurons and leaving the numerosity-only neurons, however, severely disrupted performance.

We also ran our classifier on the non-numerical visual stimulus features. Again we found that numerosity was a middling feature; numerical acuity was about average for the features tested. Clearly the stimulus representation in untrained monkeys in VIP is not specialized for numerosity. Instead, these data support the view that VIP contains a rich and generalized stimulus representation; many visual features can be extracted from neural firing rate. Furthermore, this representation is distributed, in that more information can be extracted across a broad population of unspecialized neurons than from the much smaller subset of neurons that “encode” that feature (i.e. are modulated by that feature and no other).

This view was supported by the results of the drug inactivation experiment in Chapter 5. We observed some deficits in response time following VIP inactivation as well as deficits in response time and accuracy following LIP inactivation. However, we found that the impairment was no greater in the number task than the color control task. Given the data outlined above, this is unsurprising. Many visual features are encoded in IPS, and at this point we have no positive evidence that the IPS is specialized for numerical processing in monkeys.

6.2 Future directions

There are several obvious follow up experiments that can apply the techniques developed in this dissertation. We have begun some of these, and others will hopefully be picked up soon.

In Chapter 3 we applied the stimulus space regression technique to characterize the acuity and bias of the number sense in adult human participants. One obvious application is to take a developmental perspective. Ariel Starr has already collected data comparing adults, six-year-olds, and four-year-olds on a very similar task. In general the results suggest that bias decreases and acuity increases over development. This may be particularly true of *Spacing*. Children seem to confuse more spaced-out dots with more numerous dots even more than adults do.

The stimulus space regression technique can also be applied to other neural dependent variables besides firing rate. In a study parallel to Chapter 4, Dr. Joonkoo Park collected EEG data from humans passively viewing dot array stimuli. He found that parieto-occipital electrodes are modulated by numerosity following stimulus presentation, and that this activity is strikingly early. We are also working towards applying the technique to fMRI data. Such a whole brain study could help elucidate where number is extracted in the dorsal visual stream in humans and which types of features are present in which cortical areas.

One of the main motivations for devising the stimulus space was the training study in Chapter 2. I wanted to know which aspect of the number sense training improved, bias or acuity. We have not yet replicated that study using the new orthogonalized stimulus sets. Thus, another important future direction is to examine how training influences the different model coefficients. We might also utilize different training and testing stimulus sets. It might be possible to increase

bias with certain confounded stimulus sets. Alternatively, stimulus sets in which number is highly incongruent with non-numerical features might decrease bias.

Finally, an electrophysiology follow up study would be of the utmost interest. We have provided compelling evidence that training alters the single neuron code for number. However, comparisons of numbers of neurons with certain properties across studies are inherently unreliable. We cannot rule out differences between individual animals, small differences in recording location, biases in cell selection, and other experimental vagaries. To precisely quantify how task demands alter neural encoding properties, we need within-cell comparisons. A direct comparison of encoding during passive viewing, delayed-match-to-sample, ordinal comparison, and some non-numerical comparison such as color, using the same exact stimuli would provide the critical data.

6.3 Summary

In summary, we found that the number sense is malleable into adulthood. This malleability may be reflected in the neural code for number. The number of neurons encoding number and the strength of that encoding likely increase when number is identified as the behaviorally relevant stimulus feature.

We improved our understanding of the dot array stimulus leading to several advantages. Number sense acuity can now be disentangled from the biasing effects of non-numerical features. We can also now determine which stimulus features affect a dependent variable such as neuronal firing rate.

When we applied our novel analysis to the activity we recorded in VIP, we found that there was no particular specialization of VIP for number. This finding was supported by our drug inactivation study, which failed to produce number-specific effects. Rather than specialization, we found that many features were represented. When we looked at the population as a whole, we found that many neurons played a small additive role in representing number, rather than a few neurons encoding most of the numerical information. In this sense, the VIP code for number was distributed rather than sparse.

Appendix A

Table A-1.
Stimulus features in terms of the three cardinal features, numerosity (n), *Size* (Sz), and *Spacing* (Sp), or the ratios of those features ($r_{feature}$).

Stimulus feature	Feature in terms of three cardinal features	Log of feature in terms of log of three cardinal features	Log of feature ratio in terms of log of three cardinal ratios
Total surface area (TSA)	$TSA = \sqrt{Sz \cdot n}$	$\log(TSA) = \frac{1}{2}\log(Sz) + \frac{1}{2}\log(n)$	$\log(r_{TSA}) = \frac{1}{2}\log(r_{Sz}) + \frac{1}{2}\log(r_n)$
Item surface area (ISA)	$ISA = \sqrt{\frac{Sz}{n}}$	$\log(ISA) = \frac{1}{2}\log(Sz) - \frac{1}{2}\log(n)$	$\log(r_{ISA}) = \frac{1}{2}\log(r_{Sz}) - \frac{1}{2}\log(r_n)$
Field area (FA)	$FA = \sqrt{Sp \cdot n}$	$\log(FA) = \frac{1}{2}\log(Sp) + \frac{1}{2}\log(n)$	$\log(r_{FA}) = \frac{1}{2}\log(r_{Sp}) + \frac{1}{2}\log(r_n)$
Sparsity (Spar)	$Spar = \sqrt{\frac{Sp}{n}}$	$\log(Spar) = \frac{1}{2}\log(Sp) - \frac{1}{2}\log(n)$	$\log(r_{Spar}) = \frac{1}{2}\log(r_{Sp}) - \frac{1}{2}\log(r_n)$
Total perimeter (TP)	$TP = 2\sqrt{\pi} \cdot Sz^{\frac{1}{4}} \cdot n^{\frac{3}{4}}$	$\log(\mathrm{TP}) = \log(2\sqrt{\pi}) + \frac{1}{4}\log(Sz) + \frac{3}{4}\log(n)$	$\log(r_{TP}) = \frac{1}{4}\log(r_{Sz}) + \frac{3}{4}\log(r_n)$
Item perimeter (IP)	$IP = 2\sqrt{\pi} \cdot Sz^{\frac{1}{4}} \cdot n^{-\frac{1}{4}}$	$\log(IP) = (\log(2\sqrt{\pi}) + \frac{1}{4}\log(Sz) - \frac{1}{4}\log(n)$	$\log(r_{IP}) = \frac{1}{4}\log(r_{Sz}) - \frac{1}{4}\log(r_n)$
Coverage (Cov)	$Cov = \sqrt{\frac{Sz}{Sp}}$	$\log(Cov) = \frac{1}{2}\log(Sz) - \frac{1}{2}\log(Sp)$	$\log(r_{Cov}) = \frac{1}{2}\log(r_{Sz}) - \frac{1}{2}\log(r_{Sp})$
Apparent closeness (AC)	$AC = \sqrt{Sz \cdot Sp}$	$\log(AC) = \frac{1}{2}\log(Sz) + \frac{1}{2}\log(Sp)$	$\log(r_{AP}) = \frac{1}{2}\log(r_{Sz}) + \frac{1}{2}\log(r_{Sp})$

Appendix B

B.1 ANOVA tables for Chapter 5 Experiment 1

Table B-1. ANOVA table for accuracy for monkey Br

Br VIP Accuracy

'Source'	'Sum Sq.'	'd.f.'	'Mean Sq.'	'F'	'Prob>F'
'Side'	3.447	1	3.447	22.1	0.000
'Drug'	0.186	1	0.186	1.2	0.276
'Task'	10.795	1	10.795	69.1	0.000
'Time'	0.136	1	0.136	0.9	0.351
'Side*Drug'	0.093	1	0.093	0.6	0.440
'Side*Task'	0.225	1	0.225	1.4	0.230
'Side*Time'	0.003	1	0.003	0.0	0.899
'Drug*Task'	0.000	1	0.000	0.0	0.967
'Drug*Time'	0.024	1	0.024	0.2	0.694
'Task*Time'	0.220	1	0.220	1.4	0.235
'Side*Drug*Task'	0.005	1	0.005	0.0	0.855
'Side*Drug*Time'	0.434	1	0.434	2.8	0.096
'Side*Task*Time'	0.029	1	0.029	0.2	0.669
'Drug*Task*Time'	0.230	1	0.230	1.5	0.225
'Side*Drug*Task*Time'	0.044	1	0.044	0.3	0.594
'Error'	2565.632	16425	0.156		
'Total'	2582.658	16440			

Table B-2. ANOVA table for RT for monkey Br

Br VIP RT

'Source'	'Sum Sq.'	'd.f.'	'Mean Sq.'	'F'	'Prob>F'
'Side'	3.501	1	3.501	1634.0	0.000
'Drug'	0.029	1	0.029	13.5	0.000
'Task'	0.013	1	0.013	6.0	0.015
'Time'	0.086	1	0.086	39.9	0.000
'Side*Drug'	0.291	1	0.291	135.9	0.000
'Side*Task'	0.000	1	0.000	0.0	0.961
'Side*Time'	0.000	1	0.000	0.0	0.881
'Drug*Task'	0.000	1	0.000	0.0	0.857
'Drug*Time'	0.009	1	0.009	4.3	0.038
'Task*Time'	0.002	1	0.002	0.8	0.363
'Side*Drug*Task'	0.000	1	0.000	0.0	0.982

'Side*Drug*Time'	0.002	1	0.002	1.1	0.297
'Side*Task*Time'	0.023	1	0.023	10.6	0.001
'Drug*Task*Time'	0.001	1	0.001	0.7	0.409
'Side*Drug*Task*Time'	0.002	1	0.002	0.8	0.380
'Error'	28.320	13216	0.002		
'Total'	32.400	13231			

Table B-3. ANOVA table for accuracy for monkey Sh

Sh VIP Accuracy

'Source'	'Sum Sq.'	'd.f.'	'Mean Sq.'	'F'	'Prob>F'
'Side'	0.698	1	0.698	8.4	0.004
'Drug'	0.069	1	0.069	0.8	0.363
'Task'	7.398	1	7.398	88.7	0.000
'Time'	0.117	1	0.117	1.4	0.237
'Side*Drug'	0.004	1	0.004	0.0	0.836
'Side*Task'	0.730	1	0.730	8.7	0.003
'Side*Time'	0.002	1	0.002	0.0	0.863
'Drug*Task'	0.090	1	0.090	1.1	0.299
'Drug*Time'	0.112	1	0.112	1.3	0.246
'Task*Time'	0.000	1	0.000	0.0	0.966
'Side*Drug*Task'	0.077	1	0.077	0.9	0.338
'Side*Drug*Time'	0.043	1	0.043	0.5	0.472
'Side*Task*Time'	0.402	1	0.402	4.8	0.028
'Drug*Task*Time'	0.017	1	0.017	0.2	0.648
'Side*Drug*Task*Time'	0.018	1	0.018	0.2	0.641
'Error'	1413.267	16941	0.083		
'Total'	1423.008	16956			

Table B-4. ANOVA table for RT for monkey Sh

Sh VIP RT

'Source'	'Sum Sq.'	'd.f.'	'Mean Sq.'	'F'	'Prob>F'
'Side'	0.005	1	0.005	2.0	0.159
'Drug'	0.106	1	0.106	42.6	0.000
'Task'	0.003	1	0.003	1.2	0.278
'Time'	0.341	1	0.341	137.0	0.000
'Side*Drug'	0.006	1	0.006	2.5	0.111
'Side*Task'	0.046	1	0.046	18.4	0.000
'Side*Time'	0.003	1	0.003	1.1	0.285
'Drug*Task'	0.019	1	0.019	7.5	0.006

'Drug*Time'	0.023	1	0.023	9.1	0.003
'Task*Time'	0.013	1	0.013	5.1	0.024
'Side*Drug*Task'	0.006	1	0.006	2.3	0.131
'Side*Drug*Time'	0.011	1	0.011	4.3	0.039
'Side*Task*Time'	0.005	1	0.005	1.9	0.168
'Drug*Task*Time'	0.012	1	0.012	4.9	0.028
'Side*Drug*Task*Time'	0.005	1	0.005	2.0	0.162
'Error'	38.297	15373	0.002		
'Total'	38.895	15388			

B.2 ANOVA tables for Chapter 5 Experiment 2

Table B-5. ANOVA table for accuracy for monkey Br

Br LIP Accuracy

'Source'	'Sum Sq.'	'd.f.'	'Mean Sq.'	'F'	'Prob>F'
'Side'	5.901	1	5.901	38.9	0.000
'Drug'	3.371	1	3.371	22.2	0.000
'Task'	14.267	1	14.267	94.1	0.000
'Time'	0.181	1	0.181	1.2	0.275
'Side*Drug'	2.430	1	2.430	16.0	0.000
'Side*Task'	0.108	1	0.108	0.7	0.398
'Side*Time'	0.424	1	0.424	2.8	0.095
'Drug*Task'	0.236	1	0.236	1.6	0.212
'Drug*Time'	1.502	1	1.502	9.9	0.002
'Task*Time'	1.858	1	1.858	12.3	0.000
'Side*Drug*Task'	0.070	1	0.070	0.5	0.497
'Side*Drug*Time'	2.248	1	2.248	14.8	0.000
'Side*Task*Time'	0.342	1	0.342	2.3	0.133
'Drug*Task*Time'	0.006	1	0.006	0.0	0.841
'Side*Drug*Task*Time'	0.059	1	0.059	0.4	0.531
'Error'	2459.429	16222	0.152		
'Total'	2492.685	16237			

Table B-6. ANOVA table for RT for monkey Br

Br LIP RT

'Source'	'Sum Sq.'	'd.f.'	'Mean Sq.'	'F'	'Prob>F'
'Side'	0.993	1	0.993	670.4	0.000
'Drug'	0.043	1	0.043	29.0	0.000
'Task'	0.020	1	0.020	13.2	0.000

'Time'	0.180	1	0.180	121.6	0.000
'Side*Drug'	0.464	1	0.464	313.1	0.000
'Side*Task'	0.011	1	0.011	7.6	0.006
'Side*Time'	0.001	1	0.001	0.4	0.509
'Drug*Task'	0.003	1	0.003	2.3	0.126
'Drug*Time'	0.010	1	0.010	6.7	0.009
'Task*Time'	0.000	1	0.000	0.3	0.603
'Side*Drug*Task'	0.003	1	0.003	2.0	0.155
'Side*Drug*Time'	0.004	1	0.004	2.5	0.112
'Side*Task*Time'	0.002	1	0.002	1.6	0.209
'Drug*Task*Time'	0.001	1	0.001	0.4	0.523
'Side*Drug*Task*Time'	0.000	1	0.000	0.2	0.633
'Error'	19.466	13147	0.001		
'Total'	21.219	13162			

Table B-7. ANOVA table for accuracy for monkey Sh

Sh LIP Accuracy

'Source'	'Sum Sq.'	'd.f.'	'Mean Sq.'	'F'	'Prob>F'
'Side'	1.282	1	1.282	14.7	0.000
'Drug'	0.464	1	0.464	5.3	0.021
'Task'	7.969	1	7.969	91.1	0.000
'Time'	0.165	1	0.165	1.9	0.170
'Side*Drug'	0.792	1	0.792	9.1	0.003
'Side*Task'	0.790	1	0.790	9.0	0.003
'Side*Time'	0.081	1	0.081	0.9	0.337
'Drug*Task'	0.020	1	0.020	0.2	0.631
'Drug*Time'	0.002	1	0.002	0.0	0.876
'Task*Time'	0.541	1	0.541	6.2	0.013
'Side*Drug*Task'	0.282	1	0.282	3.2	0.073
'Side*Drug*Time'	0.240	1	0.240	2.7	0.098
'Side*Task*Time'	0.016	1	0.016	0.2	0.664
'Drug*Task*Time'	0.179	1	0.179	2.1	0.152
'Side*Drug*Task*Time'	0.167	1	0.167	1.9	0.167
'Error'	1614.855	18463	0.087		
'Total'	1627.886	18478			

Table B-8. ANOVA table for RT for monkey SH

Sh LIP RT

'Source'	'Sum Sq.'	'd.f.'	'Mean Sq.'	'F'	'Prob>F'

'Side'	0.224	1	0.224	67.9	0.000
'Drug'	1.300	1	1.300	394.4	0.000
'Task'	0.003	1	0.003	1.0	0.307
'Time'	1.050	1	1.050	318.5	0.000
'Side*Drug'	0.156	1	0.156	47.5	0.000
'Side*Task'	0.058	1	0.058	17.7	0.000
'Side*Time'	0.004	1	0.004	1.1	0.297
'Drug*Task'	0.001	1	0.001	0.3	0.616
'Drug*Time'	0.188	1	0.188	57.1	0.000
'Task*Time'	0.001	1	0.001	0.3	0.614
'Side*Drug*Task'	0.000	1	0.000	0.0	0.971
'Side*Drug*Time'	0.028	1	0.028	8.6	0.003
'Side*Task*Time'	0.001	1	0.001	0.4	0.535
'Drug*Task*Time'	0.001	1	0.001	0.2	0.630
'Side*Drug*Task*Time'	0.002	1	0.002	0.7	0.410
'Error'	54.904	16659	0.003		
'Total'	57.861	16674			

References

Agrillo, Christian, Marco Dadda, and Angelo Bisazza. 2006. “Quantity Discrimination in Female Mosquitofish.” *Animal Cognition* 10 (1): 63–70. doi:10.1007/s10071-006-0036-5.

Agrillo, Christian, Laura Piffer, and Andrea Adriano. 2013. “Individual Differences in Non-Symbolic Numerical Abilities Predict Mathematical Achievements but Contradict ATOM.” *Behavioral and Brain Functions* 9 (1): 26.

Agrillo, Christian, Laura Piffer, and Angelo Bisazza. 2011. “Number versus Continuous Quantity in Numerosity Judgments by Fish.” *Cognition* 119 (2): 281–87. doi:10.1016/j.cognition.2010.10.022.

Allïk, Jüri, and Tiia Tuulmets. 1991. “Occupancy Model of Perceived Numerosity.” *Perception & Psychophysics* 49 (4): 303–14. doi:10.3758/BF03205986.

Ansari, Daniel, and Bibek Dhital. 2006. “Age-Related Changes in the Activation of the Intraparietal Sulcus during Nonsymbolic Magnitude Processing: An Event-Related Functional Magnetic Resonance Imaging Study.” *Journal of Cognitive Neuroscience* 18 (11): 1820–28. doi:10.1162/jocn.2006.18.11.1820.

Barth, Hilary C. 2008. “Judgments of Discrete and Continuous Quantity: An Illusory Stroop Effect.” *Cognition* 109 (2): 251–66. doi:10.1016/j.cognition.2008.09.002.

Bisley, James W., and Michael E. Goldberg. 2003. “Neuronal Activity in the Lateral Intraparietal Area and Spatial Attention.” *Science* 299 (5603): 81–86. doi:10.1126/science.1077395.

Bongard, Sylvia, and Andreas Nieder. 2010. “Basic Mathematical Rules Are Encoded by Primate Prefrontal Cortex Neurons.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (5): 2277–82. doi:10.1073/pnas.0909180107.

Brannon, Elizabeth M., and Herbert S. Terrace. 1998. “Ordering of the Numerosities 1 to 9 by Monkeys.” *Science* 282 (5389): 746–49. doi:10.1126/science.282.5389.746.

———. 2000. “Representation of the Numerosities 1-9 by Rhesus Macaques (Macaca Mulatta).” *Journal of Experimental Psychology. Animal Behavior Processes* 26 (1): 31–49.

Bremmer, Frank, Jean-René Duhamel, Suliann Ben Hamed, and Werner Graf. 2002. “Heading Encoding in the Macaque Ventral Intraparietal Area (VIP).” *European Journal of Neuroscience* 16 (8): 1554–68. doi:10.1046/j.1460-9568.2002.02207.x.

Bremmer, Frank, Francois Klam, Jean-Rene Duhamel, Suliann Ben Hamed, and Werner Graf. 2002. “Visual-Vestibular Interactive Responses in the Macaque Ventral Intraparietal Area (VIP).” *European Journal of Neuroscience* 16 (8): 1569–86. doi:10.1046/j.1460-9568.2002.02206.x.

Bueti, Domenica, and Vincent Walsh. 2009. “The Parietal Cortex and the Representation of Time, Space, Number and Other Magnitudes.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1525): 1831–40. doi:10.1098/rstb.2009.0028.

Burns, B., and B. E. Shepp. 1988. “Dimensional Interactions and the Structure of Psychological Space: The Representation of Hue, Saturation, and Brightness.” *Perception & Psychophysics* 43 (5): 494–507.

Cantlon, Jessica F., and Elizabeth M. Brannon. 2005. “Semantic Congruity Affects Numerical Judgments Similarly in Monkeys and Humans.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (45): 16507–11.

———. 2006. “Shared System for Ordering Small and Large Numbers in Monkeys and Humans.” *Psychological Science* 17 (5): 401–6.

———. 2007. “How Much Does Number Matter to a Monkey (Macaca Mulatta)?” *Journal of Experimental Psychology: Animal Behavior Processes* 33 (1): 32–41. doi:10.1037/0097-7403.33.1.32.

Cantlon, Jessica F., Elizabeth M. Brannon, Elizabeth J. Carter, and Kevin A. Pelphrey. 2006. “Functional Imaging of Numerical Processing in Adults and 4-Y-Old Children.” *PLoS Biology* 4 (5): e125.

Cantlon, Jessica F., Sara Cordes, Melissa E. Libertus, and Elizabeth M. Brannon. 2009. “Comment on ‘Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures.’” *Science* 323 (5910): 38b – 38b.

Cantlon, Jessica F., Michael L. Platt, and Elizabeth M. Brannon. 2009. “Beyond the Number Domain.” *Trends in Cognitive Sciences* 13 (2): 83–91. doi:10.1016/j.tics.2008.11.007.

Cantlon, Jessica F., Kelley E. Safford, and Elizabeth M. Brannon. 2010. “Spontaneous Analog Number Representations in 3-Year-Old Children.” *Developmental Science* 13 (2): 289–97. doi:10.1111/j.1467-7687.2009.00887.x.

Cappelletti, Marinella, Hilary Barth, Felipe Fregni, Elizabeth S. Spelke, and Alvaro Pascual-Leone. 2007. “rTMS over the Intraparietal Sulcus Disrupts Numerosity Processing.” *Experimental Brain Research* 179 (4): 631–42. doi:10.1007/s00221-006-0820-0.

Carey, Susan. 2011. *The Origin of Concepts*. Reprint edition. Oxford; New York: Oxford University Press.

Chen, Qi, and Tom Verguts. 2013. “Spontaneous Summation or Numerosity-Selective Coding?” *Frontiers in Human Neuroscience* 7. doi:10.3389/fnhum.2013.00886.

Chen, Qixuan, and Jingguang Li. 2014. “Association between Individual Differences in Non-Symbolic Number Acuity and Math Performance: A Meta-Analysis.” *Acta Psychologica* 148 (May): 163–72. doi:10.1016/j.actpsy.2014.01.016.

Cicchini, Guido Marco, Giovanni Anobile, and David C. Burr. 2014. “Compressive Mapping of Number to Space Reflects Dynamic Encoding Mechanisms, Not Static Logarithmic Transform.” *Proceedings of the National Academy of Sciences* 111 (21): 7867–72. doi:10.1073/pnas.1402785111.

Cipolotti, L., B. Butterworth, and G. Denes. 1991. “A Specific Deficit for Numbers in a Case of Dense Acalculia.” *Brain: A Journal of Neurology* 114 (Pt 6) (December): 2619–37.

Clearfield, Melissa W., and Kelly S. Mix. 1999. “Number Versus Contour Length in Infants’ Discrimination of Small Visual Sets.” *Psychological Science* 10 (5): 408–11. doi:10.1111/1467-9280.00177.

———. 2001. “Amount Versus Number: Infants’ Use of Area and Contour Length to Discriminate Small Sets.” *Journal of Cognition and Development* 2 (3): 243–60. doi:10.1207/S15327647JCD0203_1.

Cohen Kadosh, Roi, Nina Bien, and Alexander T. Sack. 2012. “Automatic and Intentional Number Processing Both Rely on Intact Right Parietal Cortex:

A Combined fMRI and Neuronavigated TMS Study." *Frontiers in Human Neuroscience* 6. doi:10.3389/fnhum.2012.00002.

Cohen Kadosh, Roi, and Avishai Henik. 2006. "A Common Representation for Semantic and Physical Properties: A Cognitive-Anatomical Approach." *Exp Psychol* 53: 87–94.

Cohen Kadosh, Roi, Sonja Soskic, Teresa Iuculano, Ryota Kanai, and Vincent Walsh. 2010. "Modulating Neuronal Activity Produces Specific and Long-Lasting Changes in Numerical Competence." *Current Biology* 20 (22): 2016–20. doi:10.1016/j.cub.2010.10.007.

Cohen, Laurent, Anna J. Wilson, Véronique Izard, and Stanislas Dehaene. 2009. "Acalculia and Gerstmann's Syndrome." In *The Behavioral and Cognitive Neurology of Stroke*, edited by Olivier Godefroy and Julien Bogousslavsky, 1 edition, 126–47. Cambridge: Cambridge University Press.

Colby, Carol L., Jean-René Duhamel, and Michael E. Goldberg. 1993. "Ventral Intraparietal Area of the Macaque: Anatomic Location and Visual Response Properties." *Journal of Neurophysiology* 69: 902–902.

Colby, Carol L., Jean-Rene Duhamel, and Michael E. Goldberg. 1996. "Visual, Presaccadic, and Cognitive Activation of Single Neurons in Monkey Lateral Intraparietal Area." *Journal of Neurophysiology* 76 (5): 2841–52.

Colby, Carol L., and Michael E. Goldberg. 1999. "Space and Attention in Parietal Cortex." *Annual Review of Neuroscience* 22 (1): 319–49. doi:10.1146/annurev.neuro.22.1.319.

Cordes, Sara, and Elizabeth M. Brannon. 2009. "The Relative Salience of Discrete and Continuous Quantity in Young Infants." *Developmental Science* 12 (3): 453–63. doi:10.1111/j.1467-7687.2008.00781.x.

Dakin, Steven C., Marc S. Tibber, John A. Greenwood, Frederick A. A. Kingdom, and Michael J. Morgan. 2011. "A Common Visual Metric for Approximate Number and Density." *Proceedings of the National Academy of Sciences* 108 (49): 19552–57. doi:10.1073/pnas.1113195108.

Dehaene, Stanislas. 1997. *The Number Sense How the Mind Creates Mathematics*. New York: Oxford University Press.

———. 2003. "The Neural Basis of the Weber–Fechner Law: A Logarithmic Mental Number Line." *Trends in Cognitive Sciences* 7 (4): 145–47. doi:10.1016/S1364-6613(03)00055-X.

Dehaene, Stanislas, Serge Bossini, and Pascal Giraux. 1993. “The Mental Representation of Parity and Number Magnitude.” *Journal of Experimental Psychology: General* 122 (3): 371–96. doi:10.1037/0096-3445.122.3.371.

Dehaene, Stanislas, and J P Changeux. 1993. “Development of Elementary Numerical Abilities: A Neuronal Model.” *Journal of Cognitive Neuroscience* 5 (4): 390–407. doi:10.1162/jocn.1993.5.4.390.

Dehaene, Stanislas, and Laurent Cohen. 1991. “Two Mental Calculation Systems: A Case Study of Severe Acalculia with Preserved Approximation.” *Neuropsychologia* 29 (11): 1045–54.

———. 1997. “Cerebral Pathways for Calculation: Double Dissociation between Rote Verbal and Quantitative Knowledge of Arithmetic.” *Cortex* 33 (2): 219–50. doi:10.1016/S0010-9452(08)70002-9.

Dehaene, Stanislas, Véronique Izard, and Manuela Piazza. 2005. “Control over Non-Numerical Parameters in Numerosity Experiments.” *Unpublished Manuscript (available on Www. Unicog. Org)*.

Dehaene, Stanislas, Elizabeth Spelke, Philippe Pinel, Ruxandra Stanescu, and Sanna Tsivkin. 1999. “Sources of Mathematical Thinking: Behavioral and Brain-Imaging Evidence.” *Science* 284 (5416): 970–74. doi:10.1126/science.284.5416.970.

Dehaene, Stanislas, N. Tzourio, V. Frak, L. Raynaud, L. Cohen, J. Mehler, and B. Mazoyer. 1996. “Cerebral Activations during Number Multiplication and Comparison: A PET Study.” *Neuropsychologia* 34 (11): 1097–1106.

Delazer, M., and T. Benke. 1997. “Arithmetic Facts without Meaning.” *Cortex* 33 (4): 697–710. doi:10.1016/S0010-9452(08)70727-5.

DeWind, Nicholas K, and Elizabeth M Brannon. 2012. “Malleability of the Approximate Number System: Effects of Feedback and Training.” *Frontiers in Human Neuroscience* 6: 68. doi:10.3389/fnhum.2012.00068.

Diester, Ilka, and Andreas Nieder. 2007. “Semantic Associations between Signs and Numerical Categories in the Prefrontal Cortex.” *PLoS Biol* 5 (11): e294. doi:10.1371/journal.pbio.0050294.

Dormal, Valérie, Michael Andres, and Mauro Pesenti. 2008. “Dissociation of Numerosity and Duration Processing in the Left Intraparietal Sulcus: A Transcranial Magnetic Stimulation Study.” *Cortex* 44 (4): 462–69. doi:10.1016/j.cortex.2007.08.011.

———. 2012a. “Contribution of the Right Intraparietal Sulcus to Numerosity and Length Processing: An fMRI-Guided TMS Study.” *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior* 48 (5): 623–29. doi:10.1016/j.cortex.2011.05.019.

———. 2012b. “Contribution of the Right Intraparietal Sulcus to Numerosity and Length Processing: An fMRI-Guided TMS Study.” *Cortex* 48 (5): 623–29. doi:10.1016/j.cortex.2011.05.019.

Dormal, Valérie, and Mauro Pesenti. 2009. “Common and Specific Contributions of the Intraparietal Sulci to Numerosity and Length Processing.” *Human Brain Mapping* 30 (8): 2466–76. doi:10.1002/hbm.20677.

———. 2012. “Processing Numerosity, Length and Duration in a Three-Dimensional Stroop-like Task: Towards a Gradient of Processing Automaticity?” *Psychological Research*, February. doi:10.1007/s00426-012-0414-3.

Duhamel, Jean-Rene, Carol L. Colby, and Michael E. Goldberg. 1998. “Ventral Intraparietal Area of the Macaque: Congruent Visual and Somatic Response Properties.” *Journal of Neurophysiology* 79 (1): 126–36.

Eger, Evelyn, Philipp Sterzer, Michael O Russ, Anne-Lise Giraud, and Andreas Kleinschmidt. 2003. “A Supramodal Number Representation in Human Intraparietal Cortex.” *Neuron* 37 (4): 719–25.

Evans, Theodore A., Michael J. Beran, Emily H. Harris, and Daniel F. Rice. 2009. “Quantity Judgments of Sequentially Presented Food Items by Capuchin Monkeys (Cebus Apella).” *Animal Cognition* 12 (1): 97–105. doi:10.1007/s10071-008-0174-z.

Fanini, A., and J. A. Assad. 2008. “Direction Selectivity of Neurons in the Macaque Lateral Intraparietal Area.” *Journal of Neurophysiology* 101 (1): 289–305. doi:10.1152/jn.00400.2007.

Feigenson, Lisa, Susan Carey, and Elizabeth Spelke. 2002. “Infants’ Discrimination of Number vs. Continuous Extent.” *Cognitive Psychology* 44 (1): 33–66. doi:10.1006/cogp.2001.0760.

Feigenson, Lisa, Stanislas Dehaene, and Elizabeth Spelke. 2004. “Core Systems of Number.” *Trends in Cognitive Sciences* 8 (7): 307–14. doi:10.1016/j.tics.2004.05.002.

Fias, Wim, Jan Lammertyn, Bert Reynvoet, Patrick Dupont, and Guy A. Orban. 2003. “Parietal Representation of Symbolic and Nonsymbolic Magnitude.”

Journal of Cognitive Neuroscience 15 (1): 47–56. doi:10.1162/089892903321107819.

Freedman, David J., and J. A. Assad. 2009. “Distinct Encoding of Spatial and Nonspatial Visual Information in Parietal Cortex.” *Journal of Neuroscience* 29 (17): 5671–80. doi:10.1523/JNEUROSCI.2878-08.2009.

Freedman, David J., and John A. Assad. 2006. “Experience-Dependent Representation of Visual Categories in Parietal Cortex.” *Nature* 443 (7107): 85–88. doi:10.1038/nature05078.

Freedman, David J., Maximilian Riesenhuber, Tomaso Poggio, and Earl K. Miller. 2002. “Visual Categorization and the Primate Prefrontal Cortex: Neurophysiology and Behavior.” *Journal of Neurophysiology* 88 (2): 929–41.

Frith, Christopher D., and Uta Frith. 1972. “The Solitaire Illusion: An Illusion of Numerosity.” *Perception & Psychophysics* 11 (6): 409–10.

Fuhs, Mary Wagner, and Nicole M. McNeil. 2013. “ANS Acuity and Mathematics Ability in Preschoolers from Low-Income Homes: Contributions of Inhibitory Control.” *Developmental Science* 16 (1): 136–48. doi:10.1111/desc.12013.

Gallistel, C. R., and Rochel Gelman. 1992. “Preverbal and Verbal Counting and Computation.” *Cognition*, Numerical Cognition, 44 (1–2): 43–74. doi:10.1016/0010-0277(92)90050-R.

Gebuis, Titia, and Wim Gevers. 2011. “Numerosities and Space; Indeed a Cognitive Illusion! A Reply to de Hevia and Spelke (2009).” *Cognition* 121 (2): 248–52.

Gebuis, Titia, and Bert Reynvoet. 2011. “Generating Nonsymbolic Number Stimuli.” *Behavior Research Methods* 43 (4): 981–86. doi:10.3758/s13428-011-0097-5.

———. 2012a. “The Interplay between Nonsymbolic Number and Its Continuous Visual Properties.” *Journal of Experimental Psychology: General* 141 (4): 642–48. doi:10.1037/a0026218.

———. 2012b. “Continuous Visual Properties Explain Neural Responses to Nonsymbolic Number.” *Psychophysiology* 49 (11): 1649–59. doi:10.1111/j.1469-8986.2012.01461.x.

———. 2012c. “The Role of Visual Information in Numerosity Estimation.” *PLoS ONE* 7 (5): e37426. doi:10.1371/journal.pone.0037426.

Gilmore, Camilla K., Nina Attridge, Sarah Clayton, Lucy Cragg, Samantha Johnson, Neil Marlow, Victoria Simms, and Matthew Inglis. 2013. “Individual Differences in Inhibitory Control, Not Non-Verbal Number Acuity, Correlate with Mathematics Achievement.” *PLoS ONE* 8 (6): e67374. doi:10.1371/journal.pone.0067374.

Gilmore, Camilla K., Nina Attridge, and Matthew Inglis. 2011. “Measuring the Approximate Number System.” *Quarterly Journal of Experimental Psychology (2006)* 64 (11): 2099–2109. doi:10.1080/17470218.2011.574710.

Gilmore, Camilla K., Shannon E. McCarthy, and Elizabeth S. Spelke. 2007. “Symbolic Arithmetic Knowledge without Instruction.” *Nature* 447 (7144): 589–91. doi:10.1038/nature05850.

———. 2010. “Non-Symbolic Arithmetic Abilities and Mathematics Achievement in the First Year of Formal Schooling.” *Cognition* 115 (3): 394–406. doi:10.1016/j.cognition.2010.02.002.

Ginsburg, Norman. 1976. “Effect of Item Arrangement on Perceived Numerosity: Randomness vs Regularity.” *Perceptual and Motor Skills* 42 (43): 663–68.

Goldberg, Michael E., James W. Bisley, Keith D. Powell, and Jacqueline Gottlieb. 2006. “Saccades, Salience and Attention: The Role of the Lateral Intraparietal Area in Visual Behavior.” *Progress in Brain Research* 155: 157–75. doi:10.1016/S0079-6123(06)55010-1.

Goldstone, Robert L. 1994. “Influences of Categorization on Perceptual Discrimination.” *Journal of Experimental Psychology: General* 123 (2): 178–200. doi:10.1037/0096-3445.123.2.178.

———. 1998. “Perceptual Learning.” *Annual Review of Psychology* 49 (1): 585–612. doi:10.1146/annurev.psych.49.1.585.

Graziano, Michael S. A., and Dylan F. Cooke. 2006. “Parieto-Frontal Interactions, Personal Space, and Defensive Behavior.” *Neuropsychologia* 44 (6): 845–59. doi:10.1016/j.neuropsychologia.2005.09.009.

Hahnloser, Richard H. R., Alexay A. Kozhevnikov, and Michale S. Fee. 2002. “An Ultra-Sparse Code Underlies the Generation of Neural Sequences in a Songbird.” *Nature* 419 (6902): 65–70. doi:10.1038/nature00974.

Halberda, Justin, and Lisa Feigenson. 2008a. “Developmental Change in the Acuity of the ‘Number Sense’: The Approximate Number System in 3-, 4-, 5-, and 6-Year-Olds and Adults.” *Developmental Psychology* 44 (5): 1457–65. doi:10.1037/a0012682.

———. 2008b. “Developmental Change in the Acuity of the ‘Number Sense’: The Approximate Number System in 3-, 4-, 5-, and 6-Year-Olds and Adults.” *Developmental Psychology* 44 (5): 1457–65. doi:10.1037/a0012682.

Halberda, Justin, Ryan Ly, Jeremy B Wilmer, Daniel Q Naiman, and Laura Germine. 2012. “Number Sense across the Lifespan as Revealed by a Massive Internet-Based Sample.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (28): 11116–20. doi:10.1073/pnas.1200196109.

Halberda, Justin, Michèle M M Mazzocco, and Lisa Feigenson. 2008. “Individual Differences in Non-Verbal Number Acuity Correlate with Maths Achievement.” *Nature* 455 (7213): 665–68. doi:10.1038/nature07246.

Harvey, B. M., B. P. Klein, N. Petridou, and S. O. Dumoulin. 2013. “Topographic Representation of Numerosity in the Human Parietal Cortex.” *Science* 341 (6150): 1123–26. doi:10.1126/science.1239052.

Hauser, Marc D., Fritz Tsao, Patricia Garcia, and Elizabeth S. Spelke. 2003. “Evolutionary Foundations of Number: Spontaneous Representation of Numerical Magnitudes by Cotton-Top Tamarins.” *Proceedings. Biological Sciences / The Royal Society* 270 (1523): 1441–46. doi:10.1098/rspb.2003.2414.

Holloway, Ian D, and Daniel Ansari. 2009. “Mapping Numerical Magnitudes onto Symbols: The Numerical Distance Effect and Individual Differences in Children’s Mathematics Achievement.” *Journal of Experimental Child Psychology* 103 (1): 17–29. doi:10.1016/j.jecp.2008.04.001.

Holloway, Ian D., Gavin R. Price, and Daniel Ansari. 2010. “Common and Segregated Neural Pathways for the Processing of Symbolic and Nonsymbolic Numerical Magnitude: An fMRI Study.” *NeuroImage* 49 (1): 1006–17. doi:10.1016/j.neuroimage.2009.07.071.

Honig, W. K., and Karen E. Stewart. 1989. “Discrimination of Relative Numerosity by Pigeons.” *Animal Learning & Behavior* 17 (2): 134–46.

Hubbard, Edward M., Manuela Piazza, Philippe Pinel, and Stanislas Dehaene. 2005. “Interactions between Number and Space in Parietal Cortex.” *Nature Reviews Neuroscience* 6 (6): 435–48. doi:10.1038/nrn1684.

Hurewitz, Felicia, Rochel Gelman, and Brian Schnitzer. 2006. “Sometimes Area Counts More than Number.” *Proceedings of the National Academy of Sciences* 103 (51): 19599–604. doi:10.1073/pnas.0609485103.

Hyde, Daniel C, Saeeda Khanum, and Elizabeth S Spelke. 2014. “Brief Non-Symbolic, Approximate Number Practice Enhances Subsequent Exact Symbolic Arithmetic in Children.” *Cognition* 131 (1): 92–107. doi:10.1016/j.cognition.2013.12.007.

Inglis, Matthew, and Camilla K. Gilmore. 2013. “Sampling from the Mental Number Line: How Are Approximate Number System Representations Formed?” *Cognition* 129 (1): 63–69.

Isaacs, E B, C J Edmonds, A Lucas, and D G Gadian. 2001. “Calculation Difficulties in Children of Very Low Birthweight: A Neural Correlate.” *Brain: A Journal of Neurology* 124 (Pt 9): 1701–7.

Izard, Véronique, and Stanislas Dehaene. 2008. “Calibrating the Mental Number Line.” *Cognition* 106 (3): 1221–47. doi:10.1016/j.cognition.2007.06.004.

Izard, Véronique, Coralie Sann, Elizabeth S. Spelke, and Arlette Streri. 2009. “Newborn Infants Perceive Abstract Numbers.” *Proceedings of the National Academy of Sciences* 106 (25): 10382–85. doi:10.1073/pnas.0812142106.

Jacob, Simon N., and Andreas Nieder. 2009. “Tuning to Non-Symbolic Proportions in the Human Frontoparietal Cortex.” *European Journal of Neuroscience* 30 (7): 1432–42. doi:10.1111/j.1460-9568.2009.06932.x.

Judge, Peter G., Theodore A. Evans, and Dhaval K. Vyas. 2005. “Ordinal Representation of Numeric Quantities by Brown Capuchin Monkeys (Cebus Apella).” *Journal of Experimental Psychology: Animal Behavior Processes* 31 (1): 79–94. doi:10.1037/0097-7403.31.1.79.

Kiani, R., and M. N. Shadlen. 2009. “Representation of Confidence Associated with a Decision by Neurons in the Parietal Cortex.” *Science* 324 (5928): 759–64. doi:10.1126/science.1169405.

Kilian, Annette, Sevgi Yaman, Lorenzo von Fersen, and Onur Güntürkün. 2003. “A Bottlenose Dolphin Discriminates Visual Stimuli Differing in Numerosity.” *Learning & Behavior* 31 (2): 133–42.

Kim, Jong-Nam, and Michael N. Shadlen. 1999. “Neural Correlates of a Decision in the Dorsolateral Prefrontal Cortex of the Macaque.” *Nature Neuroscience* 2 (2): 176–85. doi:10.1038/5739.

Kramer, Peter, Maria Grazia Di Bono, and Marco Zorzi. 2011. “Numerosity Estimation in Visual Stimuli in the Absence of Luminance-Based Cues.” *PLoS ONE* 6 (2): e17378. doi:10.1371/journal.pone.0017378.

Kucian, K., U. Grond, S. Rotzer, B. Henzi, C. Schönmann, F. Plangger, M. Gälli, E. Martin, and M. von Aster. 2011. “Mental Number Line Training in Children with Developmental Dyscalculia.” *NeuroImage*, Special Issue: Educational Neuroscience, 57 (3): 782–95. doi:10.1016/j.neuroimage.2011.01.070.

Lewis, Kerrie P., Sarah Jaffe, and Elizabeth M. Brannon. 2005. “Analog Number Representations in Mongoose Lemurs (Eulemur Mongoz): Evidence from a Search Task.” *Animal Cognition* 8 (4): 247–52. doi:10.1007/s10071-004-0251-x.

Libertus, Melissa E., and Elizabeth M. Brannon. 2010. “Stable Individual Differences in Number Discrimination in Infancy.” *Developmental Science* 13 (6): 900–906. doi:10.1111/j.1467-7687.2009.00948.x.

Libertus, Melissa E., Lisa Feigenson, and Justin Halberda. 2013. “Is Approximate Number Precision a Stable Predictor of Math Ability?” *Learning and Individual Differences* 25 (June): 126–33. doi:10.1016/j.lindif.2013.02.001.

Libertus, Melissa E., Ariel Starr, and Elizabeth M. Brannon. 2014. “Number Trumps Area for 7-Month-Old Infants.” *Developmental Psychology* 50 (1): 108–12. doi:10.1037/a0032986.

Libertus, Melissa E., Marty G. Woldorff, and Elizabeth M. Brannon. 2007. “Electrophysiological Evidence for Notation Independence in Numerical Processing.” *Behavioral and Brain Functions* 3 (1): 1–15.

Lipton, Jennifer S., and Elizabeth S. Spelke. 2003. *ORIGINS OF NUMBER SENSE: Large-Number Discrimination in Human Infants*.

———. 2004. “Discrimination of Large and Small Numerosities by Human Infants.” *Infancy* 5 (3): 271–90. doi:10.1207/s15327078in0503_2.

Livingstone, Margaret S., Krishna Srihasam, and Istvan A. Morocz. 2010. “The Benefit of Symbols: Monkeys Show Linear, Human-Like, Accuracy When Using Symbols to Represent Scalar Value.” *Animal Cognition* 13 (5): 711–19. doi:10.1007/s10071-010-0321-1.

Louie, Kenway, and Paul W. Glimcher. 2010. “Separating Value from Choice: Delay Discounting Activity in the Lateral Intraparietal Area.” *The Journal of*

Neuroscience 30 (16): 5498–5507. doi:10.1523/JNEUROSCI.5742-09.2010.

Lourenco, Stella F., Justin W. Bonny, Edmund P. Fernandez, and Sonia Rao. 2012. “Nonsymbolic Number and Cumulative Area Representations Contribute Shared and Unique Variance to Symbolic Math Competence.” *Proceedings of the National Academy of Sciences* 109 (46): 18737–42. doi:10.1073/pnas.1207212109.

Lourenco, Stella F., and M. R. Longo. 2010. “General Magnitude Representation in Human Infants.” *Psychological Science* 21 (6): 873–81. doi:10.1177/0956797610370158.

Lyons, Ian M., and Sian L. Beilock. 2011. “Numerical Ordering Ability Mediates the Relation between Number-Sense and Arithmetic Competence.” *Cognition* 121 (2): 256–61. doi:10.1016/j.cognition.2011.07.009.

Maloney, Erin A., Evan F. Risko, Frank Preston, Daniel Ansari, and Jonathan Fugelsang. 2010. “Challenging the Reliability and Validity of Cognitive Measures: The Case of the Numerical Distance Effect.” *Acta Psychologica* 134 (2): 154–61. doi:10.1016/j.actpsy.2010.01.006.

Matsuzawa, Tetsuro. 1985. “Use of Numbers by a Chimpanzee.” *Nature* 315 (6014): 57–59. doi:10.1038/315057a0.

Mazzocco, Michèle M. M., Lisa Feigenson, and Justin Halberda. 2011a. “Impaired Acuity of the Approximate Number System Underlies Mathematical Learning Disability (Dyscalculia).” *Child Development* 82 (4): 1224–37. doi:10.1111/j.1467-8624.2011.01608.x.

———. 2011b. “Preschoolers’ Precision of the Approximate Number System Predicts Later School Mathematics Performance.” Edited by Laurie Santos. *PLoS ONE* 6 (9): e23749. doi:10.1371/journal.pone.0023749.

Meck, Warren H., and Russell M. Church. 1983. “A Mode Control Model of Counting and Timing Processes.” *Journal of Experimental Psychology: Animal Behavior Processes* 9 (3): 320–34. doi:10.1037/0097-7403.9.3.320.

Merritt, Dustin J., Evan L. MacLean, Jeremy Chase Crawford, and Elizabeth M. Brannon. 2011. “Numerical Rule-Learning in Ring-Tailed Lemurs (Lemur Catta).” *Frontiers in Psychology* 2. doi:10.3389/fpsyg.2011.00023.

Mix, Kelly S., Janellen Huttenlocher, and Susan C. Levine. 2002. “Multiple Cues for Quantification in Infancy: Is Number One of Them?” *Psychological Bulletin* 128 (2): 278–94. doi:10.1037//0033-2909.128.2.278.

Moyer, Robert S., and Thomas K. Landauer. 1967. “Time Required for Judgements of Numerical Inequality.” *Nature* 215 (5109): 1519–20. doi:10.1038/2151519a0.

Mundy, Eleanor, and Camilla K. Gilmore. 2009. “Children’s Mapping between Symbolic and Nonsymbolic Representations of Number.” *Journal of Experimental Child Psychology*, Special Issue: Typical Development of Numerical Cognition, 103 (4): 490–502. doi:10.1016/j.jecp.2009.02.003.

Nieder, Andreas. 2012. “Supramodal Numerosity Selectivity of Neurons in Primate Prefrontal and Posterior Parietal Cortices.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (29): 11860–65. doi:10.1073/pnas.1204580109.

Nieder, Andreas, and Stanislas Dehaene. 2009. “Representation of Number in the Brain.” *Annual Review of Neuroscience* 32 (1): 185–208. doi:10.1146/annurev.neuro.051508.135550.

Nieder, Andreas, Ilka Diester, and Oana Tudusciuc. 2006. “Temporal and Spatial Enumeration Processes in the Primate Parietal Cortex.” *Science* 313 (5792): 1431–35. doi:10.1126/science.1130308.

Nieder, Andreas, David J. Freedman, and Earl K. Miller. 2002. “Representation of the Quantity of Visual Items in the Primate Prefrontal Cortex.” *Science (New York, N.Y.)* 297 (5587): 1708–11. doi:10.1126/science.1072493.

Nieder, Andreas, and K. Merten. 2007. “A Labeled-Line Code for Small and Large Numerosities in the Monkey Prefrontal Cortex.” *Journal of Neuroscience* 27 (22): 5986–93. doi:10.1523/JNEUROSCI.1056-07.2007.

Nieder, Andreas, and Earl K. Miller. 2003. “Coding of Cognitive Magnitude: Compressed Scaling of Numerical Information in the Primate Prefrontal Cortex.” *Neuron* 37 (1): 149–57. doi:10.1016/S0896-6273(02)01144-3.

———. 2004a. “A Parieto-Frontal Network for Visual Numerical Information in the Monkey.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (19): 7457–62.

———. 2004b. “Analog Numerical Representations in Rhesus Monkeys: Evidence for Parallel Processing.” *Journal of Cognitive Neuroscience* 16 (5): 889–901. doi:10.1162/089892904970807.

Odic, Darko, Howard Hock, and Justin Halberda. 2014. “Hysteresis Affects Approximate Number Discrimination in Young Children.” *Journal of Experimental Psychology: General* 143 (1): 255–65. doi:10.1037/a0030825.

Park, Joonkoo, and Elizabeth M. Brannon. 2013. “Training the Approximate Number System Improves Math Proficiency.” *Psychological Science* 24 (10): 2013–19. doi:10.1177/0956797613482944.

Pearson, John, J.D. Roitman, E.M. Brannon, M.L. Platt, and Sridhar Raghavachari. 2010. “A Physiologically-Inspired Model of Numerical Classification Based on Graded Stimulus Coding.” *Frontiers in Behavioral Neuroscience* 4 (January). doi:10.3389/neuro.08.001.2010.

Piaget, Jean. 1965. *Child’s Conception Of Number*. Copyright 1965 edition. New York: W. W. Norton & Company.

Piazza, Manuela, Andrea Facoetti, Anna Noemi Trussardi, Ilaria Berteletti, Stefano Conte, Daniela Lucangeli, Stanislas Dehaene, and Marco Zorzi. 2010. “Developmental Trajectory of Number Acuity Reveals a Severe Impairment in Developmental Dyscalculia.” *Cognition* 116 (1): 33–41. doi:10.1016/j.cognition.2010.03.012.

Piazza, Manuela, and Véronique Izard. 2009. “How Humans Count: Numerosity and the Parietal Cortex.” *The Neuroscientist* 15 (3): 261–73. doi:10.1177/1073858409333073.

Piazza, Manuela, Véronique Izard, Philippe Pinel, Denis Le Bihan, and Stanislas Dehaene. 2004. “Tuning Curves for Approximate Numerosity in the Human Intraparietal Sulcus.” *Neuron* 44 (3): 547–55.

Piazza, Manuela, Philippe Pinel, Denis Le Bihan, and Stanislas Dehaene. 2007. “A Magnitude Code Common to Numerosities and Number Symbols in Human Intraparietal Cortex.” *Neuron* 53 (2): 293–305.

Pica, Pierre, Cathy Lemer, Véronique Izard, and Stanislas Dehaene. 2004. “Exact and Approximate Arithmetic in an Amazonian Indigene Group.” *Science* 306 (5695): 499–503.

Pinel, Philippe, Stanislas Dehaene, D. Rivière, and D. LeBihan. 2001. “Modulation of Parietal Activation by Semantic Distance in a Number Comparison Task.” *NeuroImage* 14 (5): 1013–26. doi:10.1006/nimg.2001.0913.

Pinel, Philippe, Manuela Piazza, Denis Le Bihan, and Stanislas Dehaene. 2004. "Distributed and Overlapping Cerebral Representations of Number, Size, and Luminance during Comparative Judgments." *Neuron* 41 (6): 983–93. doi:10.1016/S0896-6273(04)00107-2.

Pisa, Paola Etel, and Christian Agrillo. 2009. "Quantity Discrimination in Felines: A Preliminary Investigation of the Domestic Cat (Felis Silvestris Catus)." *Journal of Ethology* 27 (2): 289–93. doi:10.1007/s10164-008-0121-0.

Platt, John R., and David M. Johnson. 1971. "Localization of Position within a Homogeneous Behavior Chain: Effects of Error Contingencies." *Learning and Motivation* 2 (4): 386–414. doi:10.1016/0023-9690(71)90020-8.

Platt, M. L., and P. W. Glimcher. 1999. "Neural Correlates of Decision Variables in Parietal Cortex." *Nature* 400 (6741): 233–38. doi:10.1038/22268.

Prather, Richard W. 2014. "Numerical Discrimination Is Mediated by Neural Coding Variation." *Cognition* 133 (3): 601–10. doi:10.1016/j.cognition.2014.08.003.

Quiroga, R Quian, L Reddy, G Kreiman, C Koch, and I Fried. 2005. "Invariant Visual Representation by Single Neurons in the Human Brain." *Nature* 435 (7045): 1102–7. doi:10.1038/nature03687.

Roberts, William A. 2010. "'Counting' Serially Presented Stimuli by Human and Nonhuman Primates and Pigeons." *Learning and Motivation*, Special Issue: Part II - In Memoriam: Stewart H. Hulse, 41 (4): 241–51. doi:10.1016/j.lmot.2010.08.002.

Roitman, Jamie D, Elizabeth M Brannon, and Michael L Platt. 2007. "Monotonic Coding of Numerosity in Macaque Lateral Intraparietal Area." *PLoS Biol* 5 (8): e208. doi:10.1371/journal.pbio.0050208.

Rose, D. 1996. "Some Reflections on (or By?) Grandmother Cells." *Perception* 25 (8): 881–86.

Rotzer, S., K. Kucian, E. Martin, M. von Aster, P. Klaver, and T. Loenneker. 2008. "Optimized Voxel-Based Morphometry in Children with Developmental Dyscalculia." *NeuroImage* 39 (1): 417–22. doi:10.1016/j.neuroimage.2007.08.045.

Rotzer, S., T. Loenneker, K. Kucian, E. Martin, P. Klaver, and M. von Aster. 2009. "Dysfunctional Neural Network of Spatial Working Memory Contributes to Developmental Dyscalculia." *Neuropsychologia* 47 (13): 2859–65. doi:10.1016/j.neuropsychologia.2009.06.009.

Rousselle, Laurence, and Marie-Pascale Noël. 2008. “The Development of Automatic Numerosity Processing in Preschoolers: Evidence for Numerosity-Perceptual Interference.” *Developmental Psychology* 44 (2): 544–60. doi:10.1037/0012-1649.44.2.544.

Sacks, Oliver W. 1985. *The Man Who Mistook His Wife for a Hat and Other Clinical Tales*. New York: Summit Books.

Sandrini, Marco, and Elena Rusconi. 2009. “A Brain for Numbers.” *Cortex* 45 (7): 796–803. doi:10.1016/j.cortex.2008.09.002.

Santens, Seppe, Chantal Roggeman, Wim Fias, and Tom Verguts. 2010. “Number Processing Pathways in Human Parietal Cortex.” *Cerebral Cortex* 20 (1): 77–88.

Santos, Laurie R., Jennifer L. Barnes, and Neha Mahajan. 2005. “Expectations about Numerical Events in Four Lemur Species (Eulemur Fulvus, Eulemur Mongoz, Lemur Catta and Varecia Rubra).” *Animal Cognition* 8 (4): 253–62. doi:10.1007/s10071-005-0252-4.

Sasanguie, Delphine, Emmy Defever, Bieke Maertens, and Bert Reynvoet. 2013. “The Approximate Number System Is Not Predictive for Symbolic Number Processing in Kindergarteners.” *Quarterly Journal of Experimental Psychology (2006)*, June. doi:10.1080/17470218.2013.803581.

Sasanguie, Delphine, Emmy Defever, Eva Van den Bussche, and Bert Reynvoet. 2011. “The Reliability of and the Relation between Non-Symbolic Numerical Distance Effects in Comparison, Same-Different Judgments and Priming.” *Acta Psychologica* 136 (1): 73–80. doi:10.1016/j.actpsy.2010.10.004.

Sasanguie, Delphine, Silke M Göbel, Kristina Moll, Karolien Smets, and Bert Reynvoet. 2013. “Approximate Number Sense, Symbolic Number Processing, or Number-Space Mappings: What Underlies Mathematics Achievement?” *Journal of Experimental Child Psychology* 114 (3): 418–31. doi:10.1016/j.jecp.2012.10.012.

Sawamura, Hiromasa, Keisetsu Shima, and Jun Tanji. 2002. “Numerical Representation for Action in the Parietal Cortex of the Monkey.” *Nature* 415 (6874): 918–22. doi:10.1038/415918a.

———. 2010. “Deficits in Action Selection Based on Numerical Information after Inactivation of the Posterior Parietal Cortex in Monkeys.” *Journal of Neurophysiology* 104 (2): 902–10. doi:10.1152/jn.01014.2009.

Scarf, D., H. Hayne, and M. Colombo. 2011. “Pigeons on Par with Primates in Numerical Competence.” *Science* 334 (6063): 1664–1664. doi:10.1126/science.1213357.

Schlack, A. 2005. “Multisensory Space Representations in the Macaque Ventral Intraparietal Area.” *Journal of Neuroscience* 25 (18): 4616–25. doi:10.1523/JNEUROSCI.0455-05.2005.

Sereno, A. B., and J. H. R. Maunsell. 1998. “Shape Selectivity in Primate Lateral Intraparietal Cortex.” *Nature* 395 (6701): 500–503. doi:10.1038/26752.

Seron, Xavier, and Mauro Pesenti. 2001. “The Number Sense Theory Needs More Empirical Evidence.” *Mind & Language* 16 (1): 76–88.

Shadlen, Michael N., and William T. Newsome. 2001. “Neural Basis of a Perceptual Decision in the Parietal Cortex (Area LIP) of the Rhesus Monkey.” *Journal of Neurophysiology* 86 (4): 1916–36.

Smirnova, A A, O F Lazareva, and Z A Zorina. 2000. “Use of Number by Crows: Investigation by Matching and Oddity Learning.” *Journal of the Experimental Analysis of Behavior* 73 (2): 163–76. doi:10.1901/jeab.2000.73-163.

Smith, Linda B., and Paul Evans. 1989. “Similarity, Identity, and Dimensions: Perceptual Classification in Children and Adults.” In *Object Perception: Structure and Process*, edited by Bryan E. Shepp and Soledad Ballesteros, 325–56. Psychology Press.

Sophian, Catherine. 2007. “Measuring Spatial Factors in Comparative Judgments about Large Numerosities.” In *Foundations of Augmented Cognition*, 157–65. Springer. http://link.springer.com/chapter/10.1007/978-3-540-73216-7_18.

Starr, Ariel, Melissa E. Libertus, and Elizabeth M. Brannon. 2013. “Number Sense in Infancy Predicts Mathematical Abilities in Childhood.” *Proceedings of the National Academy of Sciences*, October, 201302751. doi:10.1073/pnas.1302751110.

Stoianov, Ivilin, and Marco Zorzi. 2012. “Emergence of a ‘Visual Number Sense’ in Hierarchical Generative Models.” *Nature Neuroscience* 15 (2): 194–96. doi:10.1038/nn.2996.

Swaminathan, Sruthi K., Nicolas Y. Masse, and David J. Freedman. 2013. “A Comparison of Lateral and Medial Intraparietal Areas during a Visual Categorization Task.” *The Journal of Neuroscience: The Official Journal of*

the Society for Neuroscience 33 (32): 13157–70. doi:10.1523/JNEUROSCI.5723-12.2013.

Szucs, Denes, Amy Devine, Fruzsina Soltesz, Alison Nobes, and Florence Gabriel. 2013. “Developmental Dyscalculia Is Related to Visuo-Spatial Memory and Inhibition Impairment.” *Cortex* 49 (10): 2674–88. doi:10.1016/j.cortex.2013.06.007.

Szucs, Denes, Alison Nobes, Amy Devine, Florence C. Gabriel, and Titia Gebuis. 2013. “Visual Stimulus Parameters Seriously Compromise the Measurement of Approximate Number System Acuity and Comparative Effects between Adults and Children.” *Frontiers in Psychology* 4 (July). doi:10.3389/fpsyg.2013.00444.

Tibber, Marc S., Gemma S.L. Manasseh, Richard C. Clarke, Galina Gagin, Sonja N. Swanbeck, Brian Butterworth, R. Beau Lotto, and Steven C. Dakin. 2013. “Sensitivity to Numerosity Is Not a Unique Visuospatial Psychophysical Predictor of Mathematical Ability.” *Vision Research* 89 (August): 1–9. doi:10.1016/j.visres.2013.06.006.

Tokita, M., and A. Ishiguchi. 2010. “How Might the Discrepancy in the Effects of Perceptual Variables on Numerosity Judgment Be Reconciled?” *Attention, Perception & Psychophysics* 72 (7): 1839–53. doi:10.3758/APP.72.7.1839.

Tomonaga, Masaki. 2007. “Relative Numerosity Discrimination by Chimpanzees (Pan Troglodytes): Evidence for Approximate Numerical Representations.” *Animal Cognition* 11 (1): 43–57. doi:10.1007/s10071-007-0089-0.

Tomonaga, Masaki, and Tetsuro Matsuzawa. 2002a. “Enumeration of Briefly Presented Items by the Chimpanzee (Pan Troglodytes) and Humans (Homo Sapiens).” *Animal Learning & Behavior* 30 (2): 143–57.

———. 2002b. “Enumeration of Briefly Presented Items by the Chimpanzee (Pan Troglodytes) and Humans (Homo Sapiens).” *Animal Learning & Behavior* 30 (2): 143–57.

Tudusciuc, Oana, and Andreas Nieder. 2007. “Neuronal Population Coding of Continuous and Discrete Quantity in the Primate Posterior Parietal Cortex.” *Proceedings of the National Academy of Sciences* 104 (36): 14513–18. doi:10.1073/pnas.0705495104.

Vallentin, D., S. Bongard, and A. Nieder. 2012. “Numerical Rule Coding in the Prefrontal, Premotor, and Posterior Parietal Cortices of Macaques.”

Journal of Neuroscience 32 (19): 6621–30. doi:10.1523/JNEUROSCI.5071-11.2012.

Venkatraman, Vinod, Daniel Ansari, and Michael W. L. Chee. 2005. “Neural Correlates of Symbolic and Non-Symbolic Arithmetic.” *Neuropsychologia* 43 (5): 744–53. doi:10.1016/j.neuropsychologia.2004.08.005.

Verguts, Tom, and Wim Fias. 2004. “Representation of Number in Animals and Humans: A Neural Model.” *Journal of Cognitive Neuroscience* 16 (9): 1493–1504.

———. 2008. “Symbolic and Nonsymbolic Pathways of Number Processing.” *Philosophical Psychology* 21 (4): 539–54. doi:10.1080/09515080802285545.

Vinje, William E., and Jack L. Gallant. 2000. “Sparse Coding and Decorrelation in Primary Visual Cortex During Natural Vision.” *Science* 287 (5456): 1273–76. doi:10.1126/science.287.5456.1273.

Viswanathan, P., and A. Nieder. 2013. “Neuronal Correlates of a Visual ‘Sense of Number’ in Primate Parietal and Prefrontal Cortices.” *Proceedings of the National Academy of Sciences* 110 (27): 11187–92. doi:10.1073/pnas.1308141110.

Wallis, Jonathan D., Kathleen C. Anderson, and Earl K. Miller. 2001. “Single Neurons in Prefrontal Cortex Encode Abstract Rules.” *Nature* 411 (6840): 953–56. doi:10.1038/35082081.

Walsh, Vincent. 2003. “A Theory of Magnitude: Common Cortical Metrics of Time, Space and Quantity.” *Trends in Cognitive Sciences* 7 (11): 483–88.

Whalen, John, C. R. Gallistel, and Rochel Gelman. 1999. “Nonverbal Counting in Humans: The Psychophysics of Number Representation.” *Psychological Science* 10 (2): 130–37.

Wiesel, T. N., and D. H. Hubel. 1963. “SINGLE-CELL RESPONSES IN STRIATE CORTEX OF KITTENS DEPRIVED OF VISION IN ONE EYE.” *Journal of Neurophysiology* 26 (November): 1003–17.

Wilson, Anna J., Stanislas Dehaene, Philippe Pinel, Susannah K. Revkin, Laurent Cohen, and David Cohen. 2006. “Principles Underlying the Design of ‘The Number Race’, an Adaptive Computer Game for Remediation of Dyscalculia.” *Behavioral and Brain Functions: BBF* 2: 19. doi:10.1186/1744-9081-2-19.

Wilson, Anna J., Susannah K. Revkin, David Cohen, Laurent Cohen, and Stanislas Dehaene. 2006. “An Open Trial Assessment of ‘The Number Race’, an Adaptive Computer Game for Remediation of Dyscalculia.” *Behavioral and Brain Functions: BBF* 2: 20. doi:10.1186/1744-9081-2-20.

Wynn, Karen. 1998. “Psychological Foundations of Number: Numerical Competence in Human Infants.” *Trends in Cognitive Sciences* 2 (8): 296–303. doi:10.1016/S1364-6613(98)01203-0.

Xu, Fei, and Elizabeth S. Spelke. 2000. “Large Number Discrimination in 6-Month-Old Infants.” *Cognition* 74 (1): B1–11.

Xu, Fei, Elizabeth S. Spelke, and Sydney Goddard. 2005. “Number Sense in Human Infants.” *Developmental Science* 8 (1): 88–101.

Zago, Laure, Laurent Petit, Marie-Renée Turbelin, Frédéric Andersson, Mathieu Vigneau, and Nathalie Tzourio-Mazoyer. 2008. “How Verbal and Spatial Manipulation Networks Contribute to Calculation: An fMRI Study.” *Neuropsychologia* 46 (9): 2403–14. doi:10.1016/j.neuropsychologia.2008.03.001.

Zorzi, Marco, and Brian Butterworth. 1999. “A Computational Model of Number Comparison.” In *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society*, edited by Martin Hahn and Scott C. Stoness, 772–77. Lawrence Erlbaum.

Zorzi, Marco, Konstantinos Priftis, and Carlo Umiltà. 2002. “Brain Damage: Neglect Disrupts the Mental Number Line.” *Nature* 417 (6885): 138–39. doi:10.1038/417138a.

Biography

Nicholas Kurshan DeWind was born in New York City on November 29th, 1984. He graduated from Wesleyan University in 2006 with a B.A. in Neuroscience & Behavior. He is a recipient of the James B. Duke Fellowship, and the Broad Research Award for Graduate Students.

Publications

DeWind NK, Adams GK, Brannon EM, Platt ML
Modeling the approximate number system to quantify the contribution of visual stimulus features
Under review at *Cognition*.

Jones SM, Pearson J, **DeWind NK**, Paulsen D, Tenekedjieva AM, Brannon EM
Lemurs and macaques show similar numerical sensitivity
Animal Cognition, 2013 (Epub ahead of print)

McClintock SM, **DeWind NK**, Husain MM, Rowny SB, Spellman TJ, Terrace H, Lisanby SH
Disruption of Component Processes of Spatial Working Memory by Electroconvulsive Therapy but not Magnetic Seizure Therapy
International Journal of Neuropsychopharmacology, 2013. 16(1): 177-187.

DeWind NK, Brannon EM
Malleability of the approximate number system: effects of feedback and training
Frontiers Human Neuroscience, 2012, 6(68).

Merritt D, **DeWind NK**, Brannon EM
Comparative cognition of number representation
In T. Zentall & E Wasserman (Eds.), The Oxford Handbook of Comparative Cognition. Oxford, UK: Oxford University Press.